

UNIVERSIDAD AUTÓNOMA DE MADRID
FACULTAD DE PSICOLOGÍA



Tesis Doctoral

El formato del ítem y la naturaleza del constructo medido: Un estudio acerca del significado de las puntuaciones de los ítems

Autora: **Elsy Josefina Urdaneta Durán**

Directora: **Dra. María José Navas Ara**

Mayo, 2011

Pía, Miguel Alejandro y Alegría, cuanto paciencia y cuanto amor,

Ada y Rafael Ramón, tan orgullosos de mí,

a quien sino a ustedes puedo yo dedicar todo.

AGRADECIMIENTOS

A la Dra. María José Navas Ara quien dirigió este trabajo con tanto esmero, dedicación e ilusión. A ella quiero expresar mi más sentido agradecimiento porque, a pesar de la distancia geográfica, estuvo tan cerca orientando todos los asuntos relativos a la tesis, no solamente en los aspectos científicos, técnicos y administrativos, sino también aportando toda su experiencia como académica y como investigadora, llena de enseñanzas vitales indispensables para llegar a buen término con el trabajo emprendido.

A Miguel E. Viloría Ramos y Miguel A. Viloría Urdaneta por su permanente estímulo, ya fuera insistiendo en la necesidad de trabajar incansablemente o por la alegría manifestada por cada paso importante logrado.

A Alegría Viloría por su poca paciencia y persistente reclamo de la necesidad de terminar esta tarea.

A María Alejandra Urdaneta y Nelson Urdaneta, quienes se ocuparon de aligerar el trabajo con los institutos que colaboraron en la investigación y de hacer trámites engorrosos para que el trabajo saliera adelante.

A Sonia Romero y Xavier Ordoñez, acompañantes solidarios en esta travesía.

A los doctores Vicente Ponsoda y Paco Abad, quienes amablemente cargaron con el peso de los asuntos administrativos necesarios para mi estadía en Madrid.

A todos los profesores y estudiantes de los institutos Cristóbal Mendoza, Monay, Pedro José Carrillo Márquez, Coromoto, La Paz y Los Cedros, que tan cordial y desinteresadamente colaboraron en el estudio.

ÍNDICE

Presentación.....	1
Capítulo I. Aproximación teórica.....	5
1. 1. La medición de constructos psicológicos.....	7
1. 2. La fiabilidad de las puntuaciones y la validez de las inferencias.....	12
1. 3. La necesaria conjunción entre la psicología cognitiva y la psicometría.....	18
1. 4. Procesos y constructos básicos que subyacen a la ejecución del test.....	28
1. 5. Factores cognitivos que influyen en la respuesta a un ítem.....	37
1. 5. 1. Estructura del conocimiento.....	38
1. 5. 2. Habilidades cognitivas.....	39
1. 5. 3. Procesos cognitivos.....	41
1. 6. Factores no cognitivos que influyen en la respuesta a un ítem.....	44
1. 6. 1. Factores metacognitivos.....	44
1. 6. 2. Factores afectivos y conativos.....	50
1. 6. 3. Factores relativos a la tarea.....	56
1. 7. El formato del ítem.....	57
Capítulo II. Un modelo teórico para someter a prueba.....	67
2. 1. Factores cognitivos.....	72
2. 2. Factores metacognitivos.....	73
2. 3. Factores relacionados con la tarea.....	74
Capítulo III. Método.....	77
3. 1. Instrumentos.....	79
3. 1. 1. Prueba de competencias.....	79

3. 1. 2. Prueba general.....	84
3. 1. 3. Hoja de registro de los profesores.....	95
3. 2. Sujetos.....	96
3. 3. Procedimiento.....	97
3. 3. 1. Contacto con los centros escolares.....	98
3. 3. 2. Preparación de los materiales.....	99
3. 3. 3. Formación de los aplicadores.....	99
3. 3. 4. Aplicación piloto.....	100
3. 3. 5. Aplicación en el aula.....	101
3. 3. 6. Recogida de los protocolos verbales.....	103
3. 3. 7. Recogida de la información de los profesores.....	103
3. 4. Plan de análisis.....	104
3. 4. 1. Control de calidad.....	104
3. 4. 2. Análisis acerca de la estructura de los datos.....	105
3. 4. 3. Construcción y análisis preliminar de las variables predictoras.....	105
3. 4. 4. Estimación del modelo.....	109
Capítulo IV. Resultados.....	113
4. 1. Control de calidad.....	115
4. 1. 1. Estudio univariado.....	116
4. 1. 2. La no respuesta.....	121
4. 1. 3. Patrones anómalos de respuesta.....	125
4. 1.4. Calidad psicométrica.....	126
4. 1. 5. Información recabada de los profesores.....	130
4. 2. Estructura de los datos.....	131

4. 3. Examen preliminar de las variables predictoras.....	132
4. 4. Estimación y prueba del modelo.....	138
4. 4. 1. Modelo EM.....	138
4. 4. 2. Modelo RC.....	143
4. 4. 3. Modelo EA.....	147
4. 4. 4. Comparación de los modelos obtenidos en los distintos formatos....	154
Capítulo V. Discusión y conclusiones.....	157
Referencias.....	165
Anexo A.....	189
Anexo B.....	193
Anexo C.....	199
Anexo D.....	203
Anexo E.....	209
Anexo F.....	213

ÍNDICE DE TABLAS

Tabla 1. Instrumentos revisados en relación a las variables incluidas en el modelo...	85
Tabla 2. Ítems de la prueba general para medir el valor percibido de la tarea.....	89
Tabla 3. Ítems de la prueba general para medir la dificultad percibida de la tarea....	90
Tabla 4. Ítems de la prueba general para medir conocimiento autopercebido.....	90
Tabla 5. Ítems de la prueba general para medir regulación metacognitiva.....	91
Tabla 6. Ítems de la prueba general para medir conocimiento estratégico.....	92
Tabla 7. Ítems de la prueba general para medir componentes de procesamiento cognitivo.....	93
Tabla 8. Relación de centros escolares que participaron en el estudio.....	102
Tabla 9. Cuadro resumen con los ítems indicadores de cada variable predictora.....	108
Tabla 10. Porcentaje de respuestas correctas y no respondidas en cada ítem para cada grupo por instituto.....	116
Tabla 11. Distribución de frecuencias para las respuestas al ítem EM.....	117
Tabla 12. Distribución de frecuencias para las respuestas al ítem RC.....	117
Tabla 13. Distribución de frecuencias para las respuestas al ítem EA.....	117
Tabla 14. Relación entre las respuestas al ítem EM y la dificultad encontrada en el componente de procesamiento “hacer los cálculos”.....	118
Tabla 15. Distribución por sexo, edad y centro escolar de los sujetos que utilizan la estrategia visual en el ítem EM	119
Tabla 16. Descriptivos para las variables relativas al conocimiento.....	122
Tabla 17. Porcentaje de sujetos que consideraron fácil los componentes de procesamiento.....	123

Tabla 18. Porcentaje de respuestas afirmativas en los ítems correspondientes a variables del ámbito no cognitivo.....	124
Tabla 19. Parámetros e índices de fiabilidad y validez de los ítems de la prueba de competencias.....	127
Tabla 20. Fiabilidad de los subconjuntos de ítems por variable.....	129
Tabla 21. Índices para el diagnóstico del AFC para valor percibido de la tarea.....	129
Tabla 22. Índices para el diagnóstico del AFC para dificultad percibida de la tarea...	129
Tabla 23. Índices para el diagnóstico del AFC para metacognición.....	129
Tabla 24. Índices para el diagnóstico del AFC para componentes de procesamiento.....	130
Tabla 25. Valores de correlación entre conocimiento declarativo y procedimental....	131
Tabla 26. Análisis de varianza de los institutos.....	132
Tabla 27. Análisis de varianza de la forma de la prueba de competencias.....	132
Tabla 28. Estadísticos descriptivos de las variables predictoras.....	133
Tabla 29. Porcentaje de examinados que encuentran fáciles los componentes de procesamiento para los formatos EM y RC.....	133
Tabla 30. Relación entre las estrategias utilizadas al responder a los ítems de la prueba de competencias y la actuación en cada uno de ellos.....	135
Tabla 31. Relación entre la respuesta al ítem según formato y la estrategia utilizada.....	135
Tabla 32. Valores de las diferencias de medias entre formatos por variable.....	136
Tabla 33. Coeficiente de correlación entre la variable respuesta y los predictores en cada formato.....	137

Tabla 34. Resultados del modelo de regresión logística para el formato EM.....	139
Tabla 35. Resultados del modelo de regresión logística para el formato RC.....	144
Tabla 36. Resultados del modelo de regresión logística para el formato EA.....	149
Tabla 37. Correlaciones de las variables incluidas en el modelo para el formato EA.....	153
Tabla 38. Resumen de los resultados para EM, RC y EA.....	154

ÍNDICE DE FIGURAS

Figura 1. Taxonomía de constructos determinantes de las diferencias individuales.....	50
Figura 2. Modelo propuesto.....	71
Figura 3. Ítem de evaluación de la actuación.....	82
Figura 4. Ítem de respuesta corta.....	83
Figura 5. Ítem de elección múltiple.....	83
Figura 6. Distribución por sexo.....	96
Figura 7. Distribución por edad.....	97
Figura 8. Distribución por titularidad del centro.....	97
Figura 9. Proceso de modelado estadístico de los datos.....	110
Figura 10. Curva característica y función de información del ítem EM.....	127
Figura 11. Curva característica y función de información del ítem RC.....	127
Figura 12. Curva característica y función de información del ítem EA.....	127
Figura 13. Función de información del test.....	128
Figura 14. Gráficos de residuos y de influencia para el ítem EM.....	140
Figura 15. Relación entre el logit estimado y la variable CD para el formato EM.....	140
Figura 16. Relación entre el logit estimado y la variable COMB para el formato EM.....	141
Figura 17. Relación entre el logit estimado y la variable CONT para el formato EM.....	141
Figura 18. Relación entre la variable CONT y la probabilidad de respuesta correcta.....	142

Figura 19. Gráficos de residuos y de influencia para el formato RC.....	144
Figura 20. Relación entre el logit estimado y la variable CODI para el formato RC.....	146
Figura 21. Relación entre el logit estimado y la variable COMB para el formato RC.....	146
Figura 22. Gráficos de residuos y de influencia para el ítem EA.....	149
Figura 23. Relación entre el logit estimado y la variable CP para el formato RC.....	150
Figura 24. Relación entre el logit estimado y la variable COMB para el formato RC.....	150
Figura 25. Relación entre el logit estimado y la variable PLAN para el formato RC.....	151
Figura 26. Relación entre el logit estimado y la variable CONT para el formato RC.....	151
Figura 27. Relación entre el logit estimado y la variable VTF para el formato RC....	152
Figura 28. Relación entre las variables PLAN y CONT y la probabilidad de respuesta correcta al ítem con formato EA.....	153

PRESENTACIÓN

Los estudios relativos al perfeccionamiento de los instrumentos de medición psicológica y educativa ganan más importancia cada día por un variado número de razones. Hoy en día, en sociedades con un número creciente de personas con formación cada vez mejor que compiten por un número limitado de puestos de trabajo y estudio, es necesario establecer formas de selección justas, que den a cada uno lo que corresponde de manera equitativa y objetiva y con el objeto de lograr estos propósitos, los tests psicológicos y de aptitud se constituyen en herramientas de gran utilidad, tanto que en algunos países han llegado a convertirse en productos de gran demanda por sus múltiples aplicaciones.

La toma de decisiones basadas en las puntuaciones obtenidas en un test, al afectar de modo directo la vida de las personas a las que se les aplican, representa un asunto muy delicado, importante y de gran sensibilidad social, razón suficiente para que la elaboración de los mismos deba ser considerada un área de estudio que demanda rigor metodológico para avanzar en consonancia con los requerimientos tanto científicos como sociales que le exige su papel en el mundo. Un uso adecuado de tests bien contruidos y que permitan inferencias válidas es la base para una mejor toma de decisiones acerca de individuos y programas, la cual debe estar fundada en información precisa y adecuada. Así pues, la medición y la evaluación son esenciales para tomar decisiones bien fundamentadas en el área psicológica y educativa.

La evaluación de los resultados obtenidos por los estudiantes representa una fase de suma importancia en el proceso de enseñanza-aprendizaje, de manera que ésta debe hacerse con los instrumentos más confiables y apropiados. Esta es la tarea a enfrentar

para elaborar un buen test y, considerando que los ítems son la unidad básica del mismo, la construcción de un buen test pasa por escribir buenos ítems. Un buen ítem será aquel que nos permita obtener una puntuación que dé cuenta del constructo que estamos interesados en medir. Para que esto sea posible, el conjunto de ítems que conforman el test deben abarcar todos los aspectos relevantes a ese constructo. En este sentido, el formato de la pregunta que se utiliza en los exámenes es un elemento de interés relevante, pues no está claro que todos los formatos permitan evaluar cualquier constructo o, cuando menos, que lo hagan con la misma adecuación; además, pueden poner en juego procesos cognitivos diferentes al requerir del sujeto conductas distintas para responder e, igualmente, pueden no resultar igual de motivadores para todos los sujetos. Es decir, se debe tener muy en cuenta el formato de ítem a utilizar, pues como bien afirma Rodríguez (2002), esta elección lleva implícitos aspectos relacionados con el tema de la validez.

Esta es la razón por la cual la selección del formato de la pregunta ha motivado diversas investigaciones que intentan averiguar si miden lo mismo aquellas preguntas en las que el examinado debe seleccionar una respuesta entre varias o aquellas en las que el estudiante debe elaborar su respuesta, ya sea ésta una simple palabra o frase, o una respuesta larga. La revisión de estos trabajos conduce a pensar que, en general, los ítems de elección múltiple no miden lo mismo que los ítems de respuesta construida.

En esta línea, los resultados obtenidos en el estudio realizado durante el periodo de investigación tutelada ponen de manifiesto que las puntuaciones obtenidas con los distintos formatos estudiados tienen propiedades psicométricas diferentes (Urdaneta y Navas, 2007). En esta investigación se estableció la falta de equivalencia psicométrica entre los distintos formatos considerados, pero no se abordó el por qué de esta falta de equivalencia. Snow (1993) certeramente señala que la equivalencia del rasgo

establecida a través de criterios psicométricos no tiene por qué ser la misma que la establecida a través de criterios psicológicos. En tal sentido, es necesario continuar investigando en esta dirección para tratar de determinar con claridad el potencial de cada formato para medir un determinado constructo y examinar no solo la equivalencia psicométrica sino también la equivalencia psicológica de las medidas obtenidas al administrar ítems de distinto formato.

Para ello, es necesario comenzar por la definición más exacta posible del constructo a medir, determinando cuáles son los factores que intervienen en la resolución de una pregunta, ya sean variables relacionadas con los sujetos (procedentes del ámbito cognitivo pero también no cognitivo) o variables relacionadas con la tarea demandada al sujeto en dicha pregunta. Seguidamente, debe plantearse un modelo y estimar el peso de todas esas variables que podrían dar cuenta de las diferencias individuales en las puntuaciones obtenidas por los sujetos en preguntas planteadas con distinto formato. Formulado dicho modelo, habrá que ver si ese modelo ajusta adecuadamente a los ítems de distinto formato o si, por el contrario, los componentes que dan cuenta de la actuación en el ítem se expresan diferencialmente en función del formato utilizado.

Así pues, se plantea como objetivo de investigación aproximar una explicación acerca del significado psicológico de las puntuaciones obtenidas cuando el constructo es medido con ítems de distinto formato. Para investigar dónde residen las posibles diferencias según el formato utilizado para plantear la tarea, se examinarán esas variables de diversa naturaleza que, en principio, explican la actuación en el ítem y son capaces de revelar diferencias individuales en las conductas que subyacen a la puntuación obtenida por el sujeto. En definitiva, se trata de intentar determinar la contribución a la medida que pueden hacer los distintos tipos de formatos de ítem y así

establecer cuál sería el más adecuado a un determinado objetivo, grupo de sujetos, contexto de aplicación y/o constructo de interés, para producir puntajes que tengan mayor fiabilidad y poder obtener el mayor número de evidencias que respalden su validez.

Para el abordaje de esta cuestión, la investigación se estructura en cinco capítulos. El primero está constituido por una *aproximación teórica* al problema, mediante el estudio de los elementos formales dentro de los cuales se enmarca la dinámica de la medición y la naturaleza de los constructos psicológicos, para llegar a la propuesta de un *modelo teórico de ejecución del ítem* que será formulado en el segundo capítulo. En el tercer capítulo, contentivo del *método*, se detalla todo lo relacionado con la elaboración de los instrumentos utilizados, se describen los sujetos examinados, el procedimiento de recolección de la información y la forma como se hizo el análisis de datos. El cuarto capítulo contiene los *resultados* obtenidos de todos los análisis realizados. Para cerrar, el último capítulo presenta la *discusión y conclusiones*, donde se confronta el modelo teórico producto de la revisión de la literatura con su contrastación empírica, a fin de dilucidar si las puntuaciones obtenidas en cada formato están influenciadas por las distintas variables que se incluyen en el modelo teórico.

CAPÍTULO I

APROXIMACIÓN TEÓRICA

...en nuestra época actual es más difícil que nunca crear una teoría perfecta y

completa...

Kandinsky

1. 1. LA MEDICIÓN DE CONSTRUCTOS PSICOLÓGICOS

Los avances científicos se orientan hacia la mejora de la vida del hombre y de su entorno; así los científicos se plantean como finalidad el conocimiento de la realidad a objeto de poder sistematizarla y de esa manera predecir, tomar decisiones y contribuir al bienestar. Para ello, lo primero es describir los diversos objetos de estudio y para eso se necesita a su vez disponer de procedimientos de medición.

Sin entrar a profundizar en la definición que ha consumido numerosas páginas en la literatura, la medición puede entenderse como el proceso de estimar la magnitud de un atributo en el objeto estudiado asignándole un valor numérico. Esta posibilidad de cuantificación que da la medición tiene ventajas nada desdeñables. Podrían mencionarse como tales la posibilidad de descripciones y clasificaciones precisas, la formulación de hipótesis y teorías estableciendo una interrelación formal entre variables, el uso de métodos potentes de análisis matemático-estadístico para la contrastación rigurosa de estas hipótesis y teorías y la posibilidad del contraste intersubjetivo, dado que el hecho puede ser verificado de manera independiente por otros científicos (Bunge, 1989; Nunnally y Bernstein, 1995; Popper, 1963/1983). Todas estas facultades de la medición facilitan la elaboración de reconstrucciones parciales y cada vez más próximas a la realidad representadas mediante modelos.

Pero los objetos de interés del ser humano son del más variado tipo y, si bien el avance en la medición de variables de las ciencias físicas y naturales es notorio, la medición de los atributos psicológicos también ha hecho notables progresos en su corta historia, aun cuando existen importantes debates epistemológicos en cuanto a su estatus dentro de la ciencia (véase Borsboom, 2005; Borsboom y Mellemborg, 2004; Kline, 1998; Michell, 1999, 2004).

El recorrido del término medición en el campo de la ciencia no ha sido en modo alguno un camino fácil. En 1887 Helmholtz advierte que ciertos atributos físicos como masa y longitud tienen la misma estructura intrínseca matemática que los números reales positivos (aditividad y orden natural), lo que supone que la cantidad de atributo será igual a las veces que una unidad del atributo esté contenida en el objeto. Dicho de otro modo, si una variable permite la concatenación y posee relación de orden tiene entonces estructura cuantitativa, que constituye la auténtica base para la medición. En 1901 Hölder recoge y formaliza estas ideas en sus conocidos axiomas de cantidad que fijan las condiciones que se deben satisfacer para poder medir.

Estos trabajos junto con la publicación en 1903 sobre los principios matemáticos de Russell y los libros de Campbell de 1920 y 1928 crearon lo que fue llamado concepción conservadora de la medición, donde la concatenación empírica era la condición fundamental que posibilitaba la medición, sostenida por la aplicabilidad de las propiedades de la adición (Savage y Ehrlich, 1992). Sin embargo, la existencia de atributos que no eran mensurables directamente y a los que las operaciones empíricas de la concatenación no les son aplicables requerían también atención. De ahí que en la publicación de Campbell de 1928 y en la disertación de Nagel sobre los fundamentos de la medición completada en 1931 se distinguiera entre medición fundamental y medición derivada, la medición fundamental para las propiedades extensivas que tienen aditividad empírica y la medición derivada para las propiedades intensivas referidas a aquellos atributos cuyas unidades sólo pueden ser expresadas como relaciones entre las unidades de las medidas fundamentales (Savage y Ehrlich, 1992).

Durante las primeras décadas del siglo XX se produciría un intenso debate acerca de qué atributos podían medirse. Aun cuando psicólogos de la talla de Cattell y Thorndike estaban dedicados al uso práctico de la medición psicológica mediante la

elaboración y aplicación de tests mentales que llegaron a tener amplia aceptación, muchos teóricos conservadores de la medida no reconocen la medición intensiva y expresan que sin concatenación empírica no se puede hablar de medición. Este debate alcanzó un intenso auge en 1940 con el reporte final de la comisión de la *British Association for Advancement of Science* en el cual se declara que la medición fundamental es imposible para la psicología debido a que los atributos subjetivos no tienen las propiedades que se requieren para que una variable pueda ser susceptible de medición (Narens y Luce, 1986).

No obstante, esta conclusión expresada en ese informe final no apartó a los psicólogos de su intento por medir los atributos propios de su campo de estudio; al contrario, los impulsó a proponer caminos que resultaran en una medición psicológica científica posible. Stevens en sus publicaciones de los años 1946 a 1951 estableció que la medición es el proceso de asignar números a objetos o eventos de acuerdo a alguna regla y con esta definición abrió las posibilidades a la medición psicológica introduciendo los términos nominal, ordinal, intervalo y razón para referirse a particulares tipos de escalas. Partiendo de la propuesta de Stevens, un grupo de filósofos y psicólogos matemáticos investigaron las condiciones y propiedades de estas escalas para llegar a una teoría de la medición conocida como teoría representacional, detallada en los tres volúmenes del tratado *Foundations of measurement* (Krantz, Luce, Suppes y Tversky, 1971; Luce, Krantz, Suppes y Tversky, 1990; Suppes, Krantz, Luce y Tversky, 1989). Esta teoría contribuye decisivamente a ‘desbloquear’ el camino de la medición de los atributos psicológicos, al definir la escala de medida como una terna formada por un sistema relacional empírico, uno numérico y un homomorfismo o función que vincule dichos sistemas y al examinar las condiciones que ha de satisfacer un sistema relacional empírico para que exista una representación numérica del mismo.

En correspondencia con lo señalado y considerando que los métodos de la ciencia involucran equipamiento sofisticado, controles estrictos y procedimientos altamente estandarizados que tratan de garantizar la objetividad y comparabilidad de las medidas obtenidas, se ha aceptado que, aunque los tests no son lo mismo que los instrumentos de medición utilizados en las ciencias naturales, se han desarrollado importantes teorías que justifican y avalan su uso, sobre la base de la interpretación que puede hacerse de las puntuaciones obtenidas con ellos. Evidentemente los tests psicológicos han sido y siguen siendo herramientas de gran utilidad ampliamente empleadas en la medición de las variables del ámbito psicológico.

Actualmente existen tests para medir un enorme rango de capacidades, habilidades, destrezas, aptitudes, factores de personalidad, entre otros atributos psicológicos. Los padres están interesados en tener evaluaciones sobre intereses y capacidades de sus hijos, los profesores tienen que examinar a los escolares y además de evaluar su rendimiento académico deberían ser capaces de detectar problemas como dislexia o hiperactividad, los médicos deben diagnosticar estados mentales, las firmas corporativas deben seleccionar su personal mediante instrumentos que los acrediten como aptos para el puesto, los países someten a prueba poblaciones enteras de estudiantes para decidir su ingreso o no a la universidad. La diversidad de características psicológicas medidas en tales situaciones es impresionante. Las exigencias de la sociedad, que requiere clasificar, diagnosticar, evaluar, conllevan la elaboración y uso creciente de instrumentos de medición que requieren investigación y estudio cada vez más detallado. Así pues tenemos tests para usos clínicos, educativos, empresariales o para asesoramiento y orientación personal.

No es exagerado entonces afirmar que los psicólogos se enfrentan con un abrumador número de características psicológicas que deben estudiar con rigor para que

puedan ser medidas de modo aceptable, llegando a existir un igualmente aplastante número de procedimientos de prueba conocidos comúnmente como tests psicológicos.

Pero a pesar de este ingente trabajo de los psicólogos dedicados a la medición de constructos psicológicos hay un par de preguntas persistentes que quedan en el aire en relación con los tests: ¿miden realmente algo estos tests?, y si es así, ¿qué miden y qué tan bien lo miden?

La primera pregunta pertenece a un ámbito que sigue siendo problemático y que se inscribe dentro de las discusiones de carácter filosófico sobre el estatus científico de la medición psicológica que ya se mencionara previamente. A pesar del hecho de que los tests psicológicos han llegado a ser el equipamiento estándar de la investigación en las ciencias sociales y a pesar de la enorme cantidad de datos empíricos que han sido reunidos, después de un siglo de teoría e investigación sobre la medición de atributos psicológicos, investigadores como Kline (1998), Michell (1999, 2000, 2004) o Borsboom (2005) consideran que se ha prestado poca atención a la cuestión conceptual acerca de la medición en psicología y sostienen aún el debate acerca de si las puntuaciones de los tests realmente miden algo o no son más que sumas arbitrarias de respuestas a ítems. No obstante, los estudios sobre variables psicológicas como la inteligencia han mostrado un importante consenso en los hallazgos y resultados y han sido eficientes para la predicción de tal forma que, a pesar de tales consideraciones, no pueden ser rechazados tan fácilmente, aun cuando no llegaran a satisfacer a plenitud los criterios de medición científica que tales autores refieren. Es importante destacar la vigencia de este debate hoy en día, porque este hecho refleja la preocupación de los implicados en la medición psicológica en poner todo su empeño en el perfeccionamiento de las herramientas utilizadas por la psicología para la construcción de sus teorías.

La segunda pregunta ha sido objeto de más investigación en el ámbito propiamente psicológico y con resultados, de algún modo, más fructíferos y consensuados. Se identifica con el problema de la validez de las inferencias realizadas a partir de las puntuaciones en el test y de la fiabilidad de estas puntuaciones, es decir, de la calidad psicométrica de las puntuaciones obtenidas con los tests. El hecho de que se pueda extraer información que permita hacer predicciones y tomar decisiones lo más sólidamente fundadas acerca de los fenómenos que ocupan estos estudios justifica la necesidad de la investigación psicométrica, que se consolida como un área de gran interés en el contexto social y una tarea fundamental dentro de la investigación científica contemporánea. Los avances que se hagan en el perfeccionamiento de la medida de constructos psicológicos, además de dar respuestas a esas dudas planteadas en el debate sobre el estatus científico de la medición psicológica, permitirán el diseño y construcción de instrumentos con una adecuada calidad métrica que hagan posible la interpretación y utilización de las puntuaciones con el fin previsto.

1. 2. LA FIABILIDAD DE LAS PUNTUACIONES Y LA VALIDEZ DE LAS INFERENCIAS

La fiabilidad y la validez representan los dos criterios fundamentales de calidad de la medida de un atributo psicológico. Los estudios psicométricos han ido progresando en cuanto a la elaboración de estándares, técnicas y procedimientos que proporcionan los elementos que han de conducir hacia la construcción de instrumentos de medición que satisfagan estos criterios en función del objetivo de su aplicación.

La fiabilidad está asociada a la precisión de la medida obtenida. La precisión es una característica buscada en todo instrumento de medición, sea un termómetro, un reloj

o un test. Mientras mayor sea el margen de error de la medida tendremos menos confianza en ella, será una medida inestable, lo cual significa que al repetir el proceso dará valores diferentes y tendremos dudas sobre cual valor asumir como cierto. Por el contrario, un instrumento de precisión con el que se obtengan medidas con márgenes de error mínimos resulta fiable y es lo que aspiramos de un buen test. El objetivo central de los estudios de fiabilidad es la estimación de los errores aleatorios cometidos al medir los constructos psicológicos, estimación que se realiza de distinta manera según la teoría de los tests que se emplee para su estudio.

El enfoque de la Teoría Clásica de los Tests (TCT) considera el error de medida como la diferencia que existe entre la puntuación verdadera y la puntuación empírica obtenida con el test y para su estimación pone la atención en dos fuentes de error: fluctuaciones aleatorias a lo largo del tiempo y diferencias en el contenido entre los ítems del test.

En la Teoría de la Generalizabilidad (TG) también se considera el error de medida como una diferencia entre la puntuación universo y la observada, pero en el marco de esta teoría el error es descompuesto en múltiples fuentes. Esta descomposición del error considera en su análisis la variabilidad debida a los sujetos (faceta de diferenciación) y la debida a las condiciones de medida (faceta de generalización).

La Teoría de Respuesta al Ítem (TRI) utiliza un enfoque distinto en la forma de cuantificar este error. Se estima el valor del error para cada sujeto en cada nivel posible de la variable medida y se expresa mediante la inversa de la función de información del test. Desde la perspectiva de la TRI, cuanta más información proporcione el test sobre el atributo medido más fiable puede ser considerado.

Lo que hacen estas teorías para cuantificar el error aleatorio es una estimación de la variabilidad de las puntuaciones empíricas en torno a lo que cada una define como

nivel o valor real del sujeto en la característica de interés, pero utilizando perspectivas diferentes. En la TCT este error es concebido a partir de la noción de medidas replicadas y, a pesar de la existencia de distintas fuentes de error, este es incluido en un solo término sobre el que simplemente se asume que es de origen aleatorio. La TG amplía el concepto de fiabilidad al concepto de generalizabilidad, que indicará la medida en la cual una puntuación es generalizable a aquellos aspectos o facetas de interés en el estudio, lo cual significa un aporte muy interesante, ya que no sólo dará respuesta a cuestiones que caen dentro del ámbito estricto de la precisión de la medida sino que permite un acercamiento a lo relativo al significado de las puntuaciones, al hacer un examen más refinado de las fuentes de error, abordando de este modo cuestiones relativas a la adecuación de las puntuaciones. Por su parte, la función de información y el error estándar en la TRI describen la precisión de las puntuaciones obtenidas con el test como una característica que, a diferencia de lo que sucede con la TCT, varía a lo largo del continuo del rasgo latente y no como un valor único que caracteriza un instrumento de medida, dotando el concepto de fiabilidad de flexibilidad al permitir establecer la precisión que obtendremos en cada nivel del rasgo latente donde se ubique el sujeto.

Los nuevos trabajos sobre fiabilidad desde la perspectiva de la TCT y la TG muestran que lo que ha habido, sobre todo, es extensiones y refinamientos sobre el tema, más que nuevos desarrollos. Estos refinamientos tienen que ver con el gran poder y bajo costo computacional que han permitido los avances tecnológicos con el uso de ordenadores (Haertel, 2006). Cabe decir también que el planteamiento de la TRI ha permitido importantes avances técnicos en el campo de la construcción de tests, como es el caso del establecimiento de bancos de ítems que posibilitan el uso de tests adaptados al nivel del examinado en el atributo que se mide. Los progresos que se han alcanzado

en el tema de fiabilidad, gracias tanto a los desarrollos tecnológicos computacionales como a las aportaciones de la TRI, son notables y se han publicado incluso algunos estudios donde se plantean modelos que combinan la TG con la TRI (véase Haertel, 2006), mostrando como los avances desde una u otra teoría pueden ser conjugados. De especial interés resulta la publicación de Bock, Brennan y Muraki (2002), donde muestran como las estimaciones obtenidas de estudios de generalizabilidad pueden ser usadas para corregir las estimaciones de la función de información y del error estándar, en casos de existencia de dependencia condicional.

Estos avances son muy importantes y seguramente de gran utilidad para atender lo señalado en los *Standards for Educational and Psychological Testing* (AERA, APA y NCME, 1999) que establecen que “la potencia e importancia de una fuente de error particular dependerá de las condiciones específicas bajo las cuales son tomadas las medidas, de cómo son puntuadas las ejecuciones y de las interpretaciones que se hagan de estas puntuaciones” (p.26). Esto quiere decir que la fiabilidad de las medidas debe analizarse dentro del contexto de los fines para los que se hace la medición y, tal como la validez, debe ser concebida en relación al contexto y los propósitos de aplicación del test (Kane, 1996). Oportuno es entonces considerar el señalamiento de Navas (2003) quien afirma que este modo de considerar la fiabilidad “tiende a unificar los dos criterios métricos de calidad global de la prueba en un planteamiento en el que el eje rector es el uso y la interpretación que se va a hacer de las puntuaciones del test” (p.95).

La definición que hace Messick (1989) del término validez en la tercera edición del manual *Educational Measurement* señala que es un juicio evaluativo integrado del grado en el cual la evidencia empírica y los razonamientos teóricos dan soporte acerca de la adecuación y propiedad de las acciones e inferencias basadas en las puntuaciones de los tests u otros modos de evaluación. Para llegar a esta definición, la noción de

validez ha ido evolucionando a lo largo de su historia tanto desde el punto de vista de su significado como el de su tratamiento.

Se parte de una concepción empirista que considera la validez como una propiedad del test que puede ser analizada en base a su correlación con un criterio externo (Guilford, 1946; Gulliksen, 1950). Sin embargo, esta concepción de validez no era adecuada para cubrir todo lo que demandaba el ámbito escolar, en el cual desde un enfoque operacionista, las evidencias referidas al contenido son muy importantes y pueden entonces ser consideradas como fuente de argumentos de validez, sostenidos sobre la base de que los ítems del test han de ser representativos de la variable que pretende medir. Con el auge de las teorías factorialistas de la inteligencia, la validez comienza a valorarse desde una perspectiva menos ateórica en la que se le considera como el grado en que las puntuaciones del test tienen un significado psicológico en correspondencia con el constructo que se desea medir; aparece entonces en escena la validez de constructo, que será incluida como tal ya en la primera versión de los *Standards for Educational and Psychological Testing* (APA, 1954). La publicación del decisivo trabajo de Cronbach y Meehl un año más tarde servirá para apuntalar definitivamente esta nueva perspectiva, ya que va a orientar la validez hacia un marco donde el estatus teórico del atributo examinado cobra una enorme importancia.

Como puede advertirse en la definición de Messick, la evolución del concepto de validez ha ido moviéndose hacia una noción más amplia en la cual lo que se valida son las inferencias que se hagan con las medidas obtenidas con el instrumento y el proceso de recogida de evidencias de validez no se centra en el test, sino en el uso y la interpretación de las puntuaciones. El análisis de la validez se asume desde una perspectiva integradora que considera la validez de constructo como un concepto unificado que orienta el proceso de acumular el mayor número de evidencias, incluidas

las referidas al criterio o al contenido, para poder sostener la afirmación de que son válidas las inferencias que se hagan basadas en las puntuaciones.

Los dos últimas ediciones de los *Standards for Educational and Psychological Testing* (1985, 1999) recogen esta perspectiva integradora del concepto de validez y en la última edición se establecen cinco fuentes de evidencias para argumentar la validez de las inferencias, basadas en el contenido del test, los procesos de respuesta, la estructura interna del test, las relaciones con otras variables y las consecuencias del test. Señalan también que los argumentos de validez que se logren reunir deben articularse de modo tal que puedan respaldar las interpretaciones que se hagan de las puntuaciones dentro del contexto en el cual se hace uso del test, dando recomendaciones operativas específicas acerca de datos empíricos o argumentaciones teóricas que deben aportarse en determinadas situaciones.

Señala Muñoz (2004) que el estudio de la validez ha logrado más contribuciones en el campo teórico que en el campo aplicado y Borsboom (2006) incluso ironiza sobre el asunto afirmando que “la validez de constructo funciona como un agujero negro del cual nadie puede escapar: una vez que el asunto es considerado como un problema de validez de constructo, su dificultad es considerada suprahumana y su solución está más allá del conocimiento de los mortales” (p.431); añade el mismo autor que de esta forma se estará evadiendo la tarea práctica del proceso de validación y dejando sin respuesta la pregunta fundamental acerca de cuál es el significado de las puntuaciones que obtenemos de la aplicación del test. En este sentido, es interesante mencionar el estudio de Hogan y Agnello (2004) quienes hallaron que en el 94% de los trabajos que seleccionaron para su investigación se reportaba la fiabilidad y sólo en el 55% aparecía algún reporte sobre la validez, restringida ésta, en el 90% de los casos, a la correlación con otras variables. También, como señala Prieto (2004), está el hecho de que los textos

sobre la TRI no contienen apartados específicos ni propuestas particulares sobre procedimientos de validación.

Aunque resulta incuestionable que no se ha hecho un esfuerzo paralelo al realizado con la fiabilidad en el estudio de la validez, está fuera de toda duda que entender con suficiente claridad esta perspectiva unificada del concepto de validez permitirá conducir con firmeza un proceso de validación en el cual la evidencia empírica pueda ser interpretada de manera integrada con los argumentos de naturaleza teórica para dar respuesta al asunto fundamental del significado de las puntuaciones.

En definitiva, alta fiabilidad de las puntuaciones y aporte de suficientes evidencias de validez de las inferencias son las señas de identidad de un buen test y es una de las tareas a las que la psicometría debe contribuir decisivamente. Las propuestas de trabajo invitan a considerar estos criterios métricos dentro de un marco unificado, donde la fiabilidad y la validez sean examinadas a la luz de los usos y de las interpretaciones que se pretenden hacer de las puntuaciones y en el cual el proceso de validación se conciba y desarrolle como una investigación científica acerca del significado de las puntuaciones del test (Messick, 1989).

1. 3. LA NECESARIA CONJUNCIÓN ENTRE LA PSICOLOGÍA COGNITIVA Y LA PSICOMETRÍA

Para poder interpretar adecuadamente las puntuaciones de un test, no solo es preciso recoger evidencias de validez procedentes de distintas fuentes sino que es también importante contar con una teoría psicológica acerca del constructo que se desea medir. Sostienen Borsboom y Mellenberg (2007) que el planteamiento de un modelo de medida requiere enunciar una teoría sustantiva que especifique la relación causal entre

el atributo y las puntuaciones del test.

Una teoría acerca del constructo que se pretende medir brinda la posibilidad de identificar y comprender los mecanismos que operan en la mente humana cuando el examinado ejecuta una determinada tarea. La identificación pormenorizada de los elementos o factores subyacentes a la ejecución de la tarea provee un marco para la representación detallada del constructo que ayudará a identificar las fuentes de dificultad y el entramado de habilidades y destrezas concretas que debe poseer el sujeto para llevar a cabo la tarea, de modo tal que las puntuaciones obtenidas sean interpretables.

La psicología cognitiva puede ser una excelente fuente de estas teorías. La investigación cognitiva se ha ocupado del estudio de cómo la gente representa y procesa mentalmente la información y con sus hallazgos puede ser posible disponer de teorías acerca del atributo a medir que permitan comprender la ejecución del test en términos de constructos cognitivos o aprovecharlas para el mejoramiento y diseño de los tests (Bejar, 1985).

La psicología cognitiva intenta comprender y explicar los principios de la conducta inteligente a través del estudio del funcionamiento de la mente humana. Aunque los psicólogos cognitivos han trabajado desde diferentes perspectivas, tales como la de Piaget que hace énfasis en la epistemología genética, la de Vigotsky que considera el conocimiento como resultado de la interacción social, la de Chomsky que se ocupa de la psicolingüística con sus análisis acerca de las propiedades formales de las estructuras sintácticas, el enfoque del procesamiento de la información ha predominado claramente en la investigación cognitiva de las últimas décadas, contribuyendo en la formulación de teorías y procedimientos adecuados para el contraste de las mismas (Mateos, 1995).

Este enfoque está basado en la analogía entre el funcionamiento cognitivo y el funcionamiento del ordenador, considerando este funcionamiento como el proceso de manipulación formal de símbolos abstractos mediante la aplicación de reglas formales. Para desentrañar este proceso las investigaciones cognitivas proponen la existencia de una secuencia de estadios de procesamiento de la información, entre la presentación del estímulo y la emisión de la respuesta y se ocupan fundamentalmente de caracterizar cada estadio identificando los componentes específicos de las operaciones cognitivas.

En tal sentido, es posible afirmar que el énfasis de la psicología cognitiva está justamente en los procesos que se dan en la mente humana para producir la respuesta y no en el producto final que se obtiene como resultado. Del mismo modo, la psicometría, para orientar la acumulación de evidencias argumentativas de la validez, debe ser capaz de interpretar las puntuaciones obtenidas por el sujeto basándose en las conductas reveladas durante la ejecución de las tareas y no sólo o no exclusivamente en la puntuación global. De lo que se trata entonces es de aislar, describir e intentar explicar esos procesos y/o conductas para impregnar de significado psicológico las puntuaciones.

La investigación cognitiva ha puesto particular atención a una pregunta fundamental: ¿cómo piensa y aprende la gente? Se ha puesto énfasis en asuntos tales como saber de qué forma los niños desarrollan la comprensión conceptual, cómo se adquiere la experiencia en un dominio particular, cómo se organiza el conocimiento en novatos y expertos, cómo la gente razona y construye las estructuras de conocimiento, cómo se relacionan los procesos de pensamiento con las competencias, cómo se modela el conocimiento en función de contextos sociales, qué pasa en el cerebro cuando se aprende... Estos estudios se centran en la investigación de cómo se estructura mentalmente el conocimiento y de los procesos cognitivos involucrados en la

adquisición, organización y uso de ese conocimiento.

Las principales teorías que abordan el estudio de las estructuras del conocimiento distinguen entre conocimiento declarativo y conocimiento procedimental. Los estudios sobre los componentes fundamentales de la cognición utilizan la noción de arquitectura cognitiva como marco de trabajo para determinar cómo es el flujo de información y cómo es adquirida, almacenada, representada y recuperada en la mente (Anderson, 1983). Estos componentes fundamentales están sustentados, en buena medida, por la memoria de trabajo y la memoria a largo plazo, que interactúan con otros procesos psicológicos básicos como atención y percepción para formar las estructuras de conocimiento.

La evaluación de las estructuras de conocimiento y de los procesos de pensamiento requieren de tareas que revelen información acerca de patrones de pensamiento, estrategias de razonamiento e incremento del conocimiento y comprensión en el dominio evaluado (Pellegrino, Chudowsky y Glaser, 2001). Para este fin se han desarrollado algunos métodos, siendo estos del más variado tipo, desde el uso de formatos de respuesta parcialmente cerrada (Shavelson, 1972) hasta estrategias más flexibles como entrevistas en profundidad no estructuradas (Marton, 1983) y protocolos verbales (Newell y Simons, 1972). Tests como el *Architectural Registration Examination (ARE)* y el *Cisco's Net-PASS*, elaborados para evaluar conocimiento en áreas profesionales específicas o el *SRI* con tareas de aritmética son ejemplos de tests diseñados expresamente para evaluar estructuras de conocimiento en ámbitos particulares del conocimiento (Mislevy, 2006).

Los procesos cognitivos que constituyen los componentes subyacentes involucrados en la ejecución de una tarea han sido investigados por los psicólogos cognitivos a través de análisis componenciales. Estos análisis permiten detectar

diferencias entre individuos en el número y modo de uso de los componentes de procesamiento, es decir, combinación de reglas empleadas, orden de las operaciones componentes del proceso, modo de procesamiento (serial o paralelo) y tiempo requerido para ejecutar un componente (Benton y Kiewra, 1987).

El enfoque del procesamiento de la información ha seguido varias perspectivas metodológicas para la identificación y aislamiento de estos componentes. En particular cabe destacar el enfoque de los correlatos cognitivos, el de los componentes cognitivos y el de los contenidos cognitivos (Snow y Lohman, 1989; Sternberg, 1994). El enfoque de los correlatos cognitivos estudia la relación entre ejecuciones cognitivas complejas y medidas de ejecución teóricamente explícitas basadas en tareas paradigmáticas de laboratorio. El enfoque de los componentes cognitivos se orienta a probar modelos que describen como los sujetos resuelven los ítems basados en tareas del tipo actualmente encontrado en tests psicométricos estándar de habilidades mentales. El enfoque de los contenidos cognitivos hace énfasis en los conocimientos adquiridos, su organización y su uso en situaciones de prueba. Todos estos procedimientos de la investigación cognitiva están pensados para intentar explorar esa “caja negra” subyacente a la ejecución de una tarea que media entre el estímulo y la respuesta.

Los análisis de estructuras de conocimiento y procesos cognitivos para variados dominios y la identificación de los componentes subyacentes involucrados en la ejecución de una tarea son muy importantes para poder hacer una completa representación del constructo. Este hecho no ha pasado desapercibido para los psicómetras, de manera tal que algunos de estos hallazgos han sido incorporados en los desarrollos psicométricos actuales.

La aplicación de métodos de descomposición de procesos de la psicología cognitiva, que pondera el impacto de los procesos específicos en la ejecución del test,

representa un aporte significativo. En la psicología cognitiva las características de las tareas están explícitamente diseñadas para reflejar procesos, mientras que en los tests de habilidad las características de los ítems que influyen los procesos usualmente no están explícitamente especificadas, en consecuencia el impacto de los procesos específicos en la ejecución del test es desconocido (Embretson y Reise, 2000). La posibilidad de manejar información de esta naturaleza facilitará tanto el diseño de las tareas necesarias para producir las conductas que revelen el constructo como el modelado de la ejecución del ítem basado en las diversas fuentes de variación identificadas a partir de los resultados de estos análisis.

En la aproximación cognitiva la dificultad del ítem es modelada como una combinación lineal de características del ítem que son postuladas para analizar el procesamiento y no se percibe como función exclusiva del nivel de habilidad del sujeto, aun para la tarea más simple, sino que sugiere la existencia de muchas fuentes que determinan la dificultad.

Los ítems elaborados en función de principios cognitivos no solamente pueden tener valores de dificultad y discriminación adecuados a los objetivos del test sino que pueden justificarse en función de su relevancia con respecto al constructo medido en lo relativo a contenidos, estructuras y procesos cognitivos (Embretson y Gorin, 2001).

Así, la identificación del conjunto de conocimientos, habilidades y destrezas concretas que debe poseer el sujeto para llevar a cabo una tarea y de las fuentes de dificultad que este conjunto determina permitirá establecer pautas que orienten sobre cuáles deben ser los ítems y cómo debe ser el formato de los mismos para evaluar adecuadamente ese constructo.

En este sentido, los modelos de diagnóstico cognitivo pueden representar una importante contribución. Estos modelos están diseñados para medir estructuras de

conocimiento específicas y destrezas de procesamiento, con el fin de proveer información acerca de las fortalezas y debilidades del examinado (Leighton y Gierl, 2007). De la Torre (2009) señala que estos modelos permiten amalgamar los avances en las teorías psicométricas y cognitivas y pueden ser utilizados para conocer en qué medida el evaluado posee las destrezas y emplea los procesos y estrategias cognitivas necesarios para la ejecución de una tarea. De este modo, los modelos de diagnóstico cognitivo contribuyen a explicar la ejecución en el ítem a través de la descripción del nivel de habilidad del examinado (Norris, Macnab y Phillips, 2007; Rupp, Templin y Henson, 2010).

Todos estos desarrollos que posibilitan la evaluación de la dimensionalidad, la identificación de los procesos cognitivos necesarios para la ejecución de los ítems y la identificación de diferencias cualitativas entre los examinados sobre la base de la ejecución del test ya han marcado algunos avances psicométricos en lo que respecta a la formulación de algunos modelos (Cortada de Cohan, 2003; Prieto y Delgado, 1999).

Como ejemplos se pueden mencionar los modelos de TRI para medir componentes del procesamiento como los de Janssen y De Boeck (1997) y el de Maris (1995) que pusieron en práctica con tareas relativas a escritura de sinónimos, el modelo de rasgo latente lineal logístico (LLTM) de Fischer (1973), el modelo de rasgo latente multicomponente (MLTM) y el modelo de rasgo latente de componente general (GLTM), ambos de Embretson (1980, 1984) que representan modelos no compensatorios en los cuales los procesos se miden por separado en cada subtarea para poder identificar procesos cognitivos y estimar su respectiva dificultad. Por su parte, el método de atributos jerárquicos (Leighton, Gierl y Hunka, 2004) o el modelo DINA (Junker y Sitsma, 2001) representan modelos de diagnóstico cognitivo, diseñados para modelar explícitamente el conocimiento y destrezas subyacentes a la resolución de

problemas formulados en un test.

Importantes obras de referencia como *Test Design. Developments in Psychology and Psychometrics* (Embretson, 1985), *The influence of cognitive psychology on testing* (Ronning, Glover, Conoley y Witt, 1987), el manual *Educational Measurement* en sus dos últimas ediciones (Linn, 1989; Brennan, 2006), *Test theory for a new generation of tests* (Frederiksen, Mislevy y Bejar, 1993) y *Advances in educational and psychological testing: Theory and applications* (Hambleton y Zaal, 1994) ya anunciaban los importantes aportes de la psicología cognitiva a la psicometría.

Coinciden en que estos aportes están representados por la importante contribución al estudio de la validez, ya que la conceptualización de lo que representa la respuesta del examinado tiene implicaciones directas como elemento argumentativo de la validez de constructo, al ayudar a comprender el significado de las puntuaciones. Como señala Embretson (1993) el enfoque cognitivo permite abordar el tema de la validez de constructo desde una perspectiva que se preocupa por la identificación teórica del constructo.

Considerando lo señalado, puede ser sorprendente que esta conjunción que a todas luces parece necesaria e incluso inevitable entre la psicología cognitiva y la psicometría no haya llegado a un estado de consolidación pleno. Es muy probable que las distintas perspectivas metodológicas seguidas en las investigaciones cognitivas y psicométricas hayan tenido una influencia significativa en que esta conjunción no se haya cimentado.

Tradicionalmente, los estudios cognitivos se desarrollan desde la perspectiva metodológica experimental, la cual supone que los sujetos son unidades experimentales idénticas en los que las diferencias individuales deben ser controladas o neutralizadas (Prieto y Delgado, 1999). Esta perspectiva choca con la de los estudios psicométricos,

particularmente con la de las teorías de los tests, la cual se vincula estrechamente con la psicología diferencial, perspectiva que considera explícitamente diferencias entre un sujeto y otro en los atributos que mide (e.g. inteligencia, aprendizaje, aptitudes, etc).

Sin embargo, podría afirmarse que la psicología cognitiva y la psicometría se mueven hacia un punto de encuentro. Por un lado, la psicometría cada vez está más impregnada de conceptos y procedimientos propios de la psicología cognitiva, que se erige como paradigma dominante en la perspectiva psicológica de los últimos tiempos. Por el otro, la psicología cognitiva, producto de sus investigaciones, pone a disposición de la psicometría un conjunto de teorías y procedimientos que pueden ser de gran alcance para el desarrollo psicométrico (Glaser, 1981; Snow y Lohman, 1993).

Si consideramos el triángulo de la evaluación propuesto por el *National Research Council* (Pellegrino, Chudowsky y Glaser, 2001) como un marco guía para orientar el proceso de medición de atributos psicológicos, la necesidad de apoyarse en los hallazgos de la investigación cognitiva es patente. El vértice de la cognición de dicho triángulo necesita de un modelo cognitivo sobre el constructo a medir, que describa cómo se representa el conocimiento y cómo se adquiere la maestría en un determinado campo. El vértice de la observación requiere la descomposición de la ejecución de forma tal de poder establecer las tareas que podrían producir determinadas conductas, las cuales han de ser diseñadas de forma tal que eliciten tanto el conocimiento como los procesos que, según el modelo cognitivo del vértice anterior, son claves para la competencia del área en cuestión. El vértice de la interpretación precisa de un modelo capaz de incorporar toda la información aportada por las puntuaciones para explicar el constructo objeto de medida.

El test no puede ser concebido meramente como una muestra aleatoria de ítems, cuyo requisito fundamental sea cumplir los criterios tradicionales de apropiada

dificultad y alta discriminación sino como una colección estructural de ítems representativos del constructo, en el cual cada ítem enfatice las características relevantes a este constructo que está siendo medido, de modo que pueda establecerse una relación entre la estructura teórica del atributo y los patrones de respuesta al ítem observados (Lohman e Ippel, 1993; Borsboom y Mellenbergh, 2007).

Por su parte, el modelo psicométrico en el cual se apoyará la interpretación de las puntuaciones obtenidas debe ser capaz de reconocer procesos, estrategias y estructuras del conocimiento, e incluso el resto de factores que subyacen a la respuesta a un ítem. Debe tener el potencial de incorporar diversos tipos de variables en la predicción de la probabilidad de respuesta, en consecuencia, debe ser capaz de operar con constructos multidimensionales y posibilitar el tratamiento de ítems con formatos diversos, elaborados precisamente para elicitación el tipo de conductas especificadas en la definición del constructo. En pocas palabras, los modelos psicométricos deben ser idóneos para poder incorporar parámetros impregnados de significación psicológica que apoyen la validez de las inferencias realizadas a partir de las puntuaciones del test.

Los análisis cognitivos de las medidas existentes pueden ayudar a mejorar la comprensión del constructo representado, sugerir estrategias alternativas de medida y contribuir en el refinamiento de los instrumentos existentes, además de indicar nuevas direcciones, necesidades y modelos de medición (Snow y Lohman, 1989). Al construir teorías acerca de los constructos, aportar algunas técnicas que contribuyen al análisis de cómo se ejecuta una tarea y permitir el diseño de tests basados en modelos cognitivos, la psicología cognitiva puede contribuir de diversas formas a proporcionar evidencias de la validez de las inferencias realizadas a partir de las puntuaciones de los test y, de esta forma, apuntalar el significado de esas puntuaciones. Basar los tests del futuro en principios cognitivos permitirá medir los constructos psicológicos de un modo más

sofisticado y potencialmente más instructivo que en el pasado (Sternberg, 1994), al posibilitar el uso de modelos de medida que integren los hallazgos psicológicos sobre el funcionamiento cognitivo a los avances psicométricos actuales.

1. 4. PROCESOS Y CONSTRUCTOS BÁSICOS QUE SUBYACEN A LA EJECUCIÓN DEL TEST

Como se ha puesto de manifiesto, el enfoque cognitivo para la medición de constructos psicológicos sugiere que hay varios factores que determinan la respuesta al ítem. En tal sentido, para obtener medidas fiables que permitan inferencias válidas, la construcción de los tests debe estar soportada por la explicación de todos y cada uno de los procesos y estructuras que subyacen a la ejecución del test.

La meta es lograr una completa representación del constructo que facilite modelar las variables que dan cuenta de la respuesta al ítem. La naturaleza del constructo guiará la selección o construcción de tareas relevantes que permitan el desarrollo racional de rúbricas y criterios de asignación de puntuaciones basados en la mejor representación que logre hacerse del constructo, al poner de manifiesto un conjunto de factores determinantes de la ejecución.

En términos generales, puede decirse que estos factores, normalmente interactivos, son el conocimiento declarativo, el conocimiento de procedimientos y las habilidades del sujeto para llevar a cabo los procesos cognitivos que conlleva la realización de la tarea.

Pero además de todos estos factores de importancia ampliamente reconocida que indica la psicología cognitiva, es necesario considerar los aportes de investigaciones básicamente enfocadas desde la perspectiva de la psicología diferencial e incluso desde

perspectivas más integradoras que conjugan ambas orientaciones. Estos estudios revelan que existen otro tipo de variables que se perciben como influyentes en el momento de afrontar una tarea y que pueden mediar en los procesos mentales necesarios para la ejecución del ítem.

Aun cuando estas variables no son de naturaleza cognitiva, se ha señalado que pueden ejercer un efecto en la respuesta de los sujetos. Esto significa que para saber cómo se ha procedido para realizar la tarea se debería, además de tener un modelo de cómo la persona combina sus conocimientos con la información proporcionada en el ítem para dar la respuesta, contar con información relativa a las variables incluidas en esta dimensión no cognitiva, a fin de entender las diferencias individuales y facilitar una mejor comprensión de cómo otras variables que caracterizan al individuo pueden influenciar los procesos usados para resolver un problema (Hunt, 1987).

Revisando varias propuestas acerca de la estructura teórica de constructos asociados al rendimiento académico, solución de problemas e inteligencia se puede apreciar que muchos de los modelos propuestos incorporan tanto variables de orden cognitivo como variables que se ubican fuera de este ámbito.

El enfoque cognitivo propuesto por Benton y Kiewra (1987) para la evaluación del constructo rendimiento plantea la existencia de varios factores, generalmente relacionados, que contribuyen al éxito o fracaso académico. Estos factores son el conocimiento declarativo, el conocimiento de procedimientos, los procesos de control, las estrategias cognitivas y los procesos metacognitivos.

En la concepción sobre aptitud, desarrollo del aprendizaje y rendimiento, Snow (1989) identifica cinco categorías que influyen sobre cada uno de los tres constructos mencionados. Estas categorías son las estructuras conceptuales de conocimiento

declarativo, las destrezas procedimentales, las estrategias de aprendizaje, las funciones autoregulatorias y la motivación.

Nickerson (1994) identifica variados factores como componentes de la habilidad cognitiva necesaria para la solución de problemas. En su propuesta pensar y resolver problemas son tareas que abarcan un amplio rango de actividades cognitivas, a saber: operaciones o procesos básicos, conocimiento de dominios específicos, conocimiento de principios normativos del razonamiento, conocimiento de principios informales y herramientas de pensamiento relativas al uso de algoritmos y heurísticos, conocimiento metacognitivo incluyendo monitorización, control y evaluación, actitudes, disposición, valores y estilos como determinantes no cognitivos en el pensar crítico y, por último, las creencias que las personas tienen sobre su inteligencia.

La propuesta de Jackson (1994) es que, además de los constructos convencionales de habilidad cognitiva (inteligencia fluida y cristalizada), hay un número sustancial de constructos conativos que tienen importancia para el aprendizaje, de los cuales resalta el enfoque adoptado para el aprendizaje (profundo o superficial), la orientación hacia la acción vs. orientación estable, orientación hacia el dominio vs. orientación hacia la ejecución y el cuidado y la atención considerados como procesos mentales metacognitivamente guiados.

Sugrue (1995) propone un marco de trabajo en el cual sea posible, no sólo hacer especificaciones sobre los contenidos, sino también sobre los formatos utilizados partiendo de una clara conceptualización del constructo. Para la determinación de las variables a considerar en la medición del constructo se basa en tres modelos, el de Glaser, Raghavan y Baxter (1992), el de Schoenfeld (1985) y el de Smith (1991). Los componentes del primer modelo son conocimiento estructurado, representación efectiva del problema, conocimiento de procedimientos, automaticidad y destrezas

autoregulatorias. El segundo modelo se compone de recursos (conocimiento declarativo y de procedimientos), heurísticos (estrategias), control (metacognición) y sistema de creencias (acerca de sí mismo y acerca de la tarea). El tercer modelo considera factores externos e internos, pero la autora establece que los factores internos son los relevantes a la discusión sobre componentes del constructo, representados estos por afecto, experiencia, conocimiento (de hechos, conceptos, esquemas o procedimientos), heurísticos y otras características personales. Las variables contenidas en los modelos descritos son agrupadas por la autora en tres categorías que interactúan de modo complejo: estructura del conocimiento, funciones cognitivas y sistemas de creencias.

En los esfuerzos por investigar la validez de las evaluaciones en el área de ciencias Hamilton, Nussban y Snow (1997) estudiaron los procesos cognitivos involucrados en una prueba de tipo evaluación de la actuación e identificaron tres dimensiones significativamente diferentes: conocimiento y razonamiento básico, razonamiento cuantitativo y razonamiento mecánico-espacial.

El modelo propuesto por O'Neil y Schacter (1997) es una adaptación de los modelos propuestos por Baxter, Elder y Glaser (1996), Glaser, Raghavan y Baxter (1992), Mayer y Wittrock (1996) y Sugrue (1995) e incluye los siguientes cuatro componentes: comprensión de contenidos, estrategias para la solución de problemas (dependientes e independientes del dominio), metacognición y motivación.

El marco propuesto por Baxter y Glaser (1998) considera los siguientes componentes de las competencias cognitivas: representación del problema que orienta la planificación y anticipación de resultados alternativos, estrategias para alcanzar la meta que influyen la solución del problema, actividades de auto-monitorización que controlan y regulan el razonamiento y la explicación de principios que subyacen a la ejecución.

Lau y Roeser (2002), basados en las propuestas de Snow, establecen dos vías para el análisis, una relacionada con la ejecución y otra relacionada con el compromiso. Las habilidades relacionadas con la ejecución son habilidad fluida, habilidad cristalizada y habilidad espacial. La vía del compromiso, en concordancia con la teoría de la auto-eficacia de Bandura (1997) y la del valor esperado de Eccles-Parsons et al. (1983), es considerada una dimensión motivacional que abarca creencias acerca de sus propias competencias y creencias acerca del valor de la tarea.

La propuesta de Kupermintz (2002) supone ampliar el concepto de aptitud reconociendo la naturaleza dinámica y compleja de las interacciones persona-situación e incluir procesos afectivos y conativos en la explicación de las diferencias individuales en aprendizaje y rendimiento.

Shavelson, Ruiz-Primo, Li y Ayala (2003) definieron el constructo rendimiento abarcando cuatro tipos de conocimiento: declarativo, de procedimientos, esquemático y estratégico. El conocimiento declarativo incluye definiciones científicas y hechos, principalmente en forma de términos, enunciados, descripciones o datos. El conocimiento de procedimientos incluye reglas de producción del tipo si-entonces o secuencias de pasos que deben ser realizados para lograr una meta o completar una tarea. El conocimiento esquemático incluye principios, esquemas y modelos mentales usados para interpretar problemas, para investigar fallas en los sistemas y para explicar o predecir los efectos que el cambio de unos conceptos pueden tener sobre otros. El conocimiento estratégico incluye saber cuándo y cómo usar ciertos tipos de conocimiento en situaciones nuevas y emplear estrategias para reconocer las situaciones donde algunos procedimientos pueden ser llevados a cabo, para examinar las características de las tareas a fin de decidir qué conocimiento esquemático puede ser aplicado, para especificar objetivos en tareas o para controlar y monitorizar el

procesamiento cognitivo.

Ferrara et al. (2004) proponen evaluar la validez de constructo mediante el análisis de tres niveles de comprensión jerárquicamente ordenados: adquisición, procesamiento y extensión del conocimiento científico. Adquisición representa destrezas y procesos más básicos tales como observación, memorización y recuperación de datos o información. Procesamiento incluye destrezas como organización, interpretación, manipulación, verificación y resumen de datos y observaciones. Extensión representa destrezas tales como evaluación y aplicación. Cada nivel de comprensión puede ser usado en diferentes categorías, a saber: destrezas científicas y procesos cognitivos amplios. Las destrezas científicas cuyo nivel de comprensión se evalúan son: a) uso y aplicación, b) descripción y/o explicación de procedimientos científicos, c) análisis, categorización y/o formulación de hipótesis y d) creación, invención y/o aprendizaje. Los amplios procesos cognitivos son: procesamiento del lenguaje, metacognición, análisis visual-espacial, recuperación de información de la memoria a largo plazo, perspectiva y empatía.

Si se considera ahora la forma de operar en reputados estudios de evaluación educativa, se puede ver cómo el marco del programa NAEP para la evaluación del dominio Ciencias incluye una dimensión de contenidos (campos de la ciencia) y una dimensión cognitiva (saber y hacer), así como también dos dimensiones adicionales (naturaleza de la ciencia y temas relacionados) (Neidorf, Binkley y Stephens, 2006). La primera dimensión -campos de la ciencia- tiene tres áreas de contenido: ciencias de la vida, ciencias físicas y ciencias de la tierra. La segunda dimensión -la cognitiva- define los procesos y destrezas cognitivas que los ítems deben requerir de los estudiantes. Las tres categorías de esta dimensión son: comprensión de conceptos, investigación científica y razonamiento práctico.

El marco del dominio Ciencias en el estudio TIMSS 2007 (Mullis et al., 2005) está organizado en dos dimensiones, una que especifica contenidos o materias a evaluar y otra que especifica los procesos cognitivos requeridos en cada tarea. Los contenidos o materias están representados por biología, química, física y ciencias de la tierra. Los procesos cognitivos son conocimiento, aplicación y análisis. El conocimiento se refiere al conocimiento base sobre hechos de la ciencia, información, conceptos, herramientas y procedimientos. La aplicación involucra aplicaciones directas del conocimiento y la comprensión en situaciones sencillas o directas. El análisis involucra tareas más complejas relacionadas con las ciencias y requiere conocimiento y comprensión de diferentes áreas y su aplicación a situaciones nuevas.

El marco del estudio PISA propone una organización particular para cada una de las áreas que evalúa, enmarcadas en tres categorías básicas que son situación, contenido y competencias. En el área de ciencias, que se ha venido reseñando para los estudios ya señalados, la situación puede ser personal, social o global. Los contenidos de interés son conocimiento acerca de la ciencia y conocimiento del mundo natural. Las competencias son la identificación de temas científicos, la explicación científica de los fenómenos y el uso de la evidencia científica. Además, en esta área se incorpora una dimensión adicional que es la actitud ante la ciencia, la cual involucra intereses, motivaciones y creencias. Vale destacar que para el área de lectura, que representa el foco del estudio en la última edición, se incluyen dentro del marco -y con un peso hasta ahora desconocido- elementos motivacionales y metacognitivos (OECD, 2009), reconociendo que la dedicación, interés y el gusto por la lectura, así como el uso de estrategias metacognitivas pueden influenciar los niveles de destreza en los lectores.

Se observa en todo este conjunto de modelos y propuestas reseñados, además de las variables vinculadas a lo cognitivo, la consideración de otros aspectos que se

ubicarían fuera de esta dimensión. Por un lado, los referentes a los procesos metacognitivos y, por otro lado, las variables ligadas a la dimensión conativa y afectiva de la conducta del examinado.

La metacognición ha sido definida como el conocimiento y control de los propios procesos cognitivos, donde el sujeto puede seleccionar técnicas y estrategias para completar tareas con éxito. El factor metacognitivo involucra el conocimiento de las estrategias generales que pueden ser usadas en diferentes tareas, el conocimiento de las condiciones bajo las cuales esas estrategias pueden ser usadas, el conocimiento del alcance para el cual esas estrategias son efectivas y el auto-conocimiento (Pintrich, 2002).

De modo que puede afirmarse que el uso de distintos procesos metacognitivos puede ser determinante en las puntuaciones obtenidas por el examinado, pues influye en el pensamiento efectivo y la solución de problemas (Pellegrino, Chudowsky y Glaser, 2001). En este mismo orden de ideas, Sternberg (1998a) ha resaltado el uso y manejo de los componentes de control metacognitivo como una influencia importante en los procesos de respuesta del sujeto. Este autor afirma que los componentes de ejecución cognitiva son activados por los procesos o componentes de control metacognitivo y que el uso que se haga de los componentes de ejecución o procesamiento cognitivo sólo puede entenderse buscando en sus fuentes metacomponenciales.

Por otra parte, las propuestas de Snow han recalcado la importancia de considerar lo conativo y afectivo como aspectos de la mente humana que afectan tanto el aprendizaje como la ejecución de las tareas. En el modelo de aptitud propuesto por Snow y cols. (Corno, Cronbach, Kupermintz, Lohman, Mandinach, Porteus y Talbert, 2002; Snow, 1992; Snow, Corno y Jackson, 1996) se revisa y actualiza el concepto de aptitud como un repertorio multivariado de potencialidades para aprender y ejecutar

tareas, que incluyen no solamente las habilidades cognitivas sino también características afectivas y conativas. Actitudes, preferencias, valores y creencias representan factores de este tipo que se constituyen como determinantes no cognitivos de la ejecución cognitiva que hay que considerar en el modelado de la respuesta al ítem.

Pero además de las variables del ámbito cognitivo y metacognitivo y las relativas a lo conativo y afectivo, la ejecución de un ítem podría estar relacionada con el contexto y la forma en que se presenta la tarea. Un conjunto de experimentos, realizados unos desde la perspectiva cognitiva y otros desde la perspectiva diferencial, sugieren que las instrucciones del test, las condiciones de aplicación, el orden de los ítems, el formato de respuesta o el tiempo de aplicación pueden tener efectos significativos en las puntuaciones obtenidas (Embretson, 1993; Embretson y Reise, 2000; Gorin, 2006; Haertel y Wiley, 1993; Snow y Peterson, 1985).

Snow (1989) ha señalado que el patrón de influencia de las variables afectivas y conativas en el rendimiento académico varía en función de características situacionales, tal como podría ser el formato de respuesta en un escenario de evaluación. En el mismo orden de ideas, Pintrich y Schunk (2002) afirman que el tipo de tareas es una importante influencia en la motivación y en la cognición, de modo que es lógico considerar que la tarea en sí misma representa un factor que puede influir a la hora de responder a un ítem.

En atención a lo anterior, se puede afirmar que la elaboración de los tests debe estar soportada por la explicación de los procesos y estructuras del conocimiento que se desean evaluar (aspectos importantes de la validez de constructo), pero también debe considerar los factores metacognitivos, los aspectos de tipo volitivo y motivacional e incluso los aspectos contextuales relacionados con la forma de administración, formato y organización del test (Snow, 1993; Nickerson, 1994).

En definitiva, los constructores y usuarios de tests deben considerar sistemáticamente todas las variables cognitivas o no cognitivas que influenciarán en menor o mayor grado las puntuaciones obtenidas por los sujetos examinados, a fin de producir una medida lo más refinada posible que permita la toma de decisiones en condiciones de menor incertidumbre.

1. 5. FACTORES COGNITIVOS QUE INFLUYEN EN LA RESPUESTA A UN ÍTEM

La correcta y más completa definición que se pueda hacer del constructo que se desea medir es un paso fundamental en la tarea de obtener una medición de calidad. En esta tarea partimos del supuesto de que la respuesta al ítem estaría determinada, como vimos, por factores cognitivos y no cognitivos.

En las distintas propuestas revisadas acerca de la medición de constructos relacionados con la aptitud, la habilidad y el rendimiento en dominios particulares se pudo observar diversidad de modos de categorizar los componentes de la dimensión cognitiva:

- Como las categorías asociadas a la *estructura del conocimiento*, ya sea definidas como conocimiento declarativo, procedimental, e incluso el conocimiento condicional, o como conocimiento de dominios específicos, de principios normativos del razonamiento, de procedimientos y algoritmos.
- Como *habilidades cognitivas* relacionadas con la inteligencia tales como habilidad cristalizada o fluida o, más específicamente, con formas de razonamiento como por ejemplo razonamiento cuantitativo y razonamiento mecánico-espacial.

- Como los *procesos cognitivos* asociados a las tareas mentales necesarias para ejecutar el ítem.

1. 5. 1. Estructura del conocimiento

Entre los trabajos más destacados sobre la estructura del conocimiento están los desarrollados por Anderson (1983, 1993) con la llamada teoría de la arquitectura cognitiva (ACT), en la que propone estructuras de conocimiento declarativo y procedimental. Anderson ofrece un marco teórico que incluye la codificación inicial de la información, su almacenamiento y recuperación como conocimiento declarativo o procedimental. Se asocia el conocimiento declarativo principalmente al proceso cognitivo de memorización, sin obviar la recuperación y los procesos de activación para la búsqueda en la memoria. El conocimiento procedimental, que representa el conocimiento de cómo hacer las cosas, está asociado al concepto de producción. La producción, concepto de la ACT, constituye la estructura de almacenamiento del conocimiento procedimental y provee la conexión entre conocimiento declarativo y conducta. McNamara (1994) señala que el conocimiento declarativo puede ser verbalizado, visualizado o declarado de alguna manera y el conocimiento procedimental soporta destrezas cognitivas tales como razonamiento, solución de problemas y comprensión del lenguaje.

El conocimiento declarativo es representado en términos de una estructura de datos que consiste en un número limitado de elementos con relaciones específicas expresados mediante representaciones analógicas o simbólicas. En el lenguaje de los sistemas de producción las unidades en las que se representa el conocimiento en la memoria de trabajo son llamadas elementos de la memoria de trabajo. Estos elementos asumen roles específicos, con una estructura que puede ser lineal o jerárquica y pueden

ser codificados con respecto a su posición en el espacio o en función de sus relaciones semánticas.

El conocimiento procedimental se almacena en forma de producciones que pueden concebirse como reglas de condición-acción que establecen la acción que hay que realizar y las condiciones en que debe llevarse a cabo. Los sistemas de producción son las redes en las cuales se organizan las producciones.

Otras maneras de clasificar los componentes cognitivos en función del tipo de conocimiento son las presentadas por Sugrue (1993, 1995) o por Nickerson (1994), relacionadas específicamente con la solución de problemas y que pueden encajar en las categorías generales de estructura del conocimiento. En el modelo de Sugrue, comprensión de conceptos, de principios y de relaciones que vinculan estos conceptos y principios puede ser considerado conocimiento declarativo, así como asociación de conceptos y principios a condiciones y procedimientos que requieran su aplicación puede asociarse a conocimiento procedimental. De la misma manera, los factores presentados por Nickerson, a saber, conocimiento de dominios específicos, conocimiento de principios normativos del razonamiento y conocimiento de algoritmos y heurísticos pueden ser distribuidos dentro de las categorías asociadas a conocimiento declarativo y procedimental.

1. 5. 2. Habilidades cognitivas

Las categorías asociadas a las habilidades cognitivas provienen directamente de las teorías y estudios acerca de la inteligencia.

Por lo que respecta a la teoría, en la actualidad se considera el modelo Cattell-Horn-Carroll (CHC) un marco de trabajo de gran utilidad y mayoritariamente aceptado en el estudio de la inteligencia y habilidades cognitivas. Esta teoría compendia las

teorías Gf-Gc de Cattell y Horn (Cattell, 1963; Horn, 1968) y la teoría de los tres estratos de Carroll (1993), proporcionando una taxonomía muy completa y soportada empíricamente. Tanto la teoría Gf Gc como la teoría de los tres estratos son modelos de naturaleza jerárquica. La diferencia fundamental está en que el modelo Gf-Gc de Cattell y Horn incluye solo dos niveles jerárquicos (uno con habilidades cognitivas amplias y otro con las habilidades cognitivas específicas), mientras que el modelo de Carroll incluye un tercer estrato donde está el factor *g*.

La teoría CHC es también un modelo jerárquico con 3 estratos. En el más amplio (estrato III), está el factor *g* o inteligencia general. En el siguiente (estrato II) estarían los 10 grandes factores de la inteligencia, con la inteligencia fluida (Gf) y cristalizada (Gc) a la cabeza junto con memoria a corto plazo (Gsm), percepción visual (Gv), percepción auditiva (Ga), habilidad para recuperar información de la memoria a largo plazo (Glr), velocidad de procesamiento cognitivo (Gs), habilidades para lectura y escritura (Grw), razonamiento cuantitativo (Gq) y tiempo de reacción (Gt). Cada uno de estos grandes factores se subdividen en factores más específicos (por ejemplo, procesos inductivos, vocabulario, memoria visual, visualización, relaciones espaciales, razonamiento secuencial, entre otros) que constituyen el último estrato (Estrato I) del modelo (McGrew, 1997), que se configura casi como una tabla periódica de elementos de la inteligencia.

Por lo que respecta a la investigación, ésta se orienta principalmente al estudio de diferencias entre novatos y expertos y a lo relativo a la inteligencia y a las habilidades para realizar determinados tipos de tareas cognitivas (Bennett, Morley, Quardt, Singley et al., 1998; Carroll, 1993; Lohman, 1993).

Pese a la importancia de estas teorías de la inteligencia, es importante señalar la observación de Carroll (1993) apuntando a que cualquier análisis del conjunto de tareas

implicadas en la ejecución de tests para medir rendimiento en dominios específicos debería inevitablemente distribuirse en conocimiento declarativo y procedimental, puesto que las diferencias individuales en habilidades cognitivas hacen referencia a diferencias en el conocimiento declarativo y procedimental que los sujetos hayan adquirido y podido demostrar en la tarea en cuestión. Afirma además, que mejor que tratar de dibujar una línea divisoria entre habilidades cognitivas y rendimiento será precisar un continuo que vaya desde habilidades más generales a los más especializados tipos de conocimiento.

1. 5. 3. Procesos cognitivos

En relación a esta categoría se han presentado interesantes propuestas, básicamente relacionadas con el estudio de la inteligencia y la solución de problemas.

Newel y Simons (1972), en su texto *Human problem solving*, establecieron un conjunto de fases por las que pasa la mente humana para la solución de problemas. De modo general estas fases serían:

- Un proceso inicial que produce dentro del solucionador una representación interna del ambiente externo, al mismo tiempo que selecciona el espacio problema.
- Una vez que el problema es representado internamente, el sistema responde a la selección de un método particular para solucionar el problema. Un método es un proceso que brinda alguna relación racional para obtener la solución del problema en términos de la representación interna.
- El método seleccionado es aplicado, pero puede pasar que, como resultado de los procesos de monitorización, la ejecución del método sea detenida.

- Cuando termina la aplicación del método, se abren tres opciones al solucionador: (a) intentar otro método, (b) seleccionar una representación interna diferente y reformular el problema, o (c) abandonar el intento por resolver el problema.
- Durante esta operación, un método puede producir nuevos problemas (e. g. sub-metas) y el sujeto puede intentar resolver alguna de estas sub-metas. Debe tenerse en cuenta que el sistema de procesamiento de la información trabaja de forma serial, es decir, el solucionador puede ver varias cosas a la vez pero solo hace una cosa al momento.
- Una vez generada la solución el solucionador para y reporta el resultado.

Desde una perspectiva razonablemente más sencilla, Bourne et al. (1979) proponen tres etapas para la solución del problema:

- Preparación, en la cual debe producirse la comprensión de la tarea a realizar y tiene como resultado la construcción de una representación del problema
- Producción, en la cual se generan posibles soluciones y debe dar como resultado una solución potencial
- Juicio, que permite la evaluación de la solución generada y dará como resultado una decisión que juzgue si el problema está resuelto o si se necesita más trabajo.

Por su parte, Sternberg (1985, 1998b) ha propuesto en su teoría triárquica de la inteligencia tres subteorías: la contextual que relaciona la inteligencia con el mundo externo del individuo, la componencial que relaciona la inteligencia con el mundo interior del individuo y la experiencial que relaciona la inteligencia con la experiencia como elemento mediador del mundo internos y externo del individuo. De particular interés en este punto son los componentes especificados por la subteoría componencial,

representados por los metacomponentes, los componentes de ejecución y los componentes de adquisición de conocimiento.

Los componentes de ejecución son los que se emplean en la realización de la tarea y varían en función de la naturaleza de la misma, por lo que su número puede ser elevado. En este sentido, es aconsejable que puedan ser organizados según las fases más generales de ejecución de la tarea. De modo amplio, Sternberg (1985) ha establecido las siguientes fases:

- Codificación, en la que se produce la percepción inicial y el almacenamiento de información nueva.
- Combinación y comparación, en la que el sujeto se ocupa de reunir, organizar y comparar información para buscar la solución. Pueden utilizarse en esta fase componentes tales como ubicar, inferir, aplicar, calcular, justificar, entre otros.
- Respuesta, en la cual el sujeto comunica la solución obtenida.

Es interesante señalar que estas dos últimas propuestas resultan bastante similares al agrupar la ejecución de la tarea en tres fases. Sin embargo, para Sternberg, el funcionamiento de los componentes de ejecución es comprensible sólo si se logra entender cómo son activados por los componentes de control ejecutivo, es decir, presumiendo siempre que la ejecución es el producto de la acción de un variado tipo de componentes, establecidos en su teoría triárquica.

En la taxonomía de Bloom, revisada por Krathwohl (2002), una dimensión está relacionada con una jerarquía de procesos cognitivos (recordar, comprender, aplicar, analizar, evaluar y crear) y la otra con el conocimiento o contenido. Como hemos visto en un epígrafe anterior, los estudios NAEP, PISA y TIMSS operan con un marco de trabajo de esas características, si bien con su propia definición tanto de los procesos como de los contenidos.

La consideración de todos estos elementos del ámbito cognitivo como determinantes de la ejecución de una tarea es un aspecto de importancia fundamental para la representación del constructo. No obstante, como ya hemos mencionado, todos estos factores estarán influidos en mayor o menor grado por otras variables cuya naturaleza no es estrictamente cognitiva y que van a ser abordados en el siguiente apartado.

1. 6. FACTORES NO COGNITIVOS QUE INFLUYEN EN LA RESPUESTA A UN ÍTEM

El conjunto de factores que se encuentran fuera del ámbito cognitivo y que se supone determinan en alguna medida la ejecución del ítem es bastante amplio. Para su estudio pueden subdividirse en categorías, que si bien no conforman grupos estancos, permiten organizar el campo:

- Factores metacognitivos.
- Factores afectivos y conativos.
- Factores relativos a la tarea.

1. 6. 1. Factores metacognitivos

La metacognición se refiere al conocimiento y control del propio sistema cognitivo (Brown, 1987). Este conocimiento guía los procesos de aprendizaje y de solución de problemas, mejorando la eficiencia de las conductas orientadas hacia el logro de la meta (Davidson, Deuser y Sternberg, 1994). En su teoría triárquica de la inteligencia Sternberg (1985) establece que la ejecución de la tarea está influenciada por los metacomponentes, que permiten al examinado coordinar el uso de su conocimiento y le facilitan un repertorio de estrategias para lograr una meta, es decir, son empleados

para una efectiva ejecución de la tarea.

Existe una diversidad de modos de considerar los determinantes metacognitivos. Brown (1987) incluso señala la existencia de fuentes de confusión para conceptualizar la metacognición. Una primera fuente estaría representada en la dificultad que a menudo se presenta para distinguir entre lo metacognitivo y lo cognitivo, pues las funciones cognitivas y metacognitivas pueden resultar intercambiables. Por ejemplo, hacerse preguntas sobre la lectura para revisar lo aprendido puede verse como una estrategia cognitiva (memorizar y comprender) o como una actividad metacognitiva (función de monitorización). Una segunda fuente de confusión se da al considerar como metacognición únicamente lo que está relacionado con el conocimiento acerca de la cognición o sólo lo que concierne a la regulación de la cognición. Por tanto, es necesario aclarar que consideraremos metacognición como un aspecto global que encierra tanto el conocimiento de la cognición, como los procesos de control usados en la ejecución.

1. 6. 1. 1. Metacognición como conocimiento de la cognición

El conocimiento metacognitivo es conocimiento de la cognición en general, así como también conciencia y conocimiento de nuestra propia cognición (Krathwohl, 2002). Incluye el conocimiento de estrategias que pueden ser usadas en diferentes tareas (conocimiento estratégico), el conocimiento de condiciones bajo las cuales el uso de una determinada estrategia es útil (conocimiento condicional) y el conocimiento de sí mismo (conocimiento auto-percibido, denominado por algunos autores experiencia metacognitiva).

Conocimiento estratégico

Es el conocimiento de estrategias generales tanto para el aprendizaje como para la solución de problemas. Se relaciona con el conocimiento de estrategias usadas para

lograr de modo más efectivo un resultado satisfactorio y son de aplicación general a todas las disciplinas académicas.

Las estrategias metacognitivas para el aprendizaje pueden ser organizadas en tres categorías generales: estrategias de práctica, de elaboración y de organización. Estas estrategias son usadas por los estudiantes para memorizar material, extraer el significado de textos y comprender sus apuntes de clases, libros y otros materiales.

Las estrategias metacognitivas para la solución de problemas están representadas básicamente por los heurísticos, que son estrategias generales que pueden ser usadas en variados dominios. Entre los principales heurísticos están el análisis medio-fin que evalúa la meta y los recursos que se tienen para alcanzarla, el trabajo hacia atrás (working backward) útil cuando conocemos el resultado al que queremos llegar, el trabajo hacia adelante (working forward) en el cual se trabaja directamente hacia la meta y supone una profunda comprensión del problema, la descomposición del problema en submetas para lo cual el problema se subdivide en una secuencia de tareas más simples para resolver el problema combinando las soluciones de las tareas más simples, la consideración de problemas análogos que permitan relacionar el problema con otro resuelto anteriormente y el uso de estrategias mixtas (Bruning, Schraw, Norbu y Ronning, 2005; Fan y Zhu, 2007; Nickerson, 1994).

Conocimiento condicional

Es el conocimiento desarrollado por el sujeto acerca de condiciones y tareas en que el uso de determinadas estrategias es adecuado (Paris, Lipson y Wixson, 1983; Pintrich, 2002). Se relaciona con el conocimiento que se utiliza para lograr la meta de un modo más eficiente, el cual permite valorar si la información que se tiene sobre la tarea es suficiente o pobre, si está bien estructurada o desorganizada, si la tarea resulta

familiar o no. En suma, consiste en saber cuándo y por qué usar el conocimiento declarativo y procedimental.

Conocimiento auto-percibido o experiencia metacognitiva

La experiencia metacognitiva está constituida por la percepción que tienen los sujetos acerca de sus propios recursos ante una tarea, por ejemplo, sentir que no se tiene suficiente información para llegar a la meta o sentir que se conoce el problema. Estas experiencias pueden ser breves o de mayor duración, y simples o complejas en su contenido (Flavell, 1992). Las experiencias metacognitivas pueden actuar como activadores de estrategias cognitivas y de funciones de regulación metacognitiva. Es una forma de auto-conocimiento que incluye el conocimiento de nuestras fortalezas y debilidades, del alcance y profundidad de nuestro conocimiento, implica cualquier cosa que la persona pueda creer acerca de su naturaleza y la de otras personas como procesadores cognitivos. Como se verá más adelante, esta definición presenta analogías con otras como competencia percibida o auto-eficacia, que han sido estudiadas dentro de las teorías motivacionales (Eccles et al., 1983; Wigfield, 1994; Eccles y Wigfield, 1995, 2002; Bandura, 1993, 1997), pero considerada como percepción sobre nuestro propio conocimiento puede incluirse en la categoría metacognitiva.

1. 6. 1. 2. Metacognición como procesos de control

Los procesos de control son funciones metacognitivas que sirven como los medios a través de los cuales los estudiantes autogestionan sus conductas de aprendizaje y de solución de problemas. Los procesos de control, también denominados regulación metacognitiva, implican la habilidad para controlar el uso de actividades cognitivas involucradas en la ejecución de tareas. Comprende la selección de un plan de trabajo adecuado, la monitorización de la ejecución del plan y la evaluación de su efectividad (Justice y Dornan, 2001). También se refieren al modo como los sujetos pueden

controlar su propia memoria, aprendizaje y otras actividades cognitivas (Sperling et al., 2004).

Existen muchas formas de categorizar los procesos metacognitivos involucrados en la regulación. Davidson, Deuser y Sternberg (1994) señalan cuatro procesos metacognitivos que son identificación del problema, representación, planificación de cómo proceder y evaluación de la solución. En su investigación sobre el impacto del formato del ítem en la actividad metacognitiva, O'Neil y Brown (1998) denominan estrategias cognitivas las que se corresponden con las actividades metacognitivas utilizadas en la planificación y representación del problema y auto-inspección a las actividades de monitorización y evaluación. Schraw y Dennison (1994) establecen cinco categorías del factor regulación de la cognición, a saber: planificación, manejo de información, monitorización, depuración y evaluación. Sperling y cols. (2004) afirman que la regulación de la cognición, teóricamente, contiene diferentes sub-componentes que incluyen planificación, selección, monitorización, evaluación y depuración. En la aplicación de regulación metacognitiva a problemas de matemática, Schoenfeld (1987) distingue entre análisis, exploración y verificación. El marco de la taxonomía sobre funciones metacognitivas de Meijer, Veenman y van Hout-Wolters (2006) contiene seis categorías: orientación, planificación, ejecución, monitorización, evaluación y elaboración.

Como se desprende de lo anterior, los componentes de la actividad metacognitiva pueden distinguirse a varios niveles de especificidad. No obstante, en un grueso grupo de investigaciones se señala que a un nivel alto se presentan planificación, monitorización y evaluación como las principales funciones de control ejecutivo (Brown, 1987; Dowson y McInerney, 1998, 2004; García y McKeachie, 2005; Justice y Dornan, 2001; Pintrich y DeGroot, 1990; Pintrich y cols., 1993).

Planificación

Está definida como el esquema mental previo acerca de cómo el sujeto abordará la tarea. Se refiere a la implementación de estrategias organizacionales concebidas para asegurar el logro exitoso de la meta e implica una esmerada preparación para completar la tarea. En el contexto del aprendizaje esta actividad conlleva programar, establecer metas realistas y ordenar apropiadamente el ambiente de trabajo y en el contexto de la solución de problemas se relaciona con el empleo de tácticas que faciliten la comprensión del problema, la organización de la información y la selección de las estrategias de trabajo. De modo general, la planificación también implica priorizar y manejar el tiempo adecuadamente (Davidson, Deuser y Sternberg, 1994; Dowson y McInerney, 2004).

Monitorización

Está definida como el control que se tiene sobre los aspectos de la ejecución tales como el tiempo, el esfuerzo, limitaciones del problema y se refiere a la implementación de medidas de auto-chequeo. En el contexto del aprendizaje implica los intentos sistemáticos de evaluar la asimilación y organización de la información aprendida y en el contexto de la solución de problemas da información acerca del estado de los sistemas cognitivos y de la actividad cognitiva dirigida a la consecución de la meta: organización, progreso, éxito y resultado (Dowson y McInerney, 2004; Kluwe, 1987).

Evaluación

Se relaciona con la valoración que se obtiene del proceso de monitorización y permite decidir si es necesario adaptar las estrategias dirigidas a contrarrestar dificultades o deficiencias identificadas durante el monitoreo. En el contexto del aprendizaje las estrategias específicas de esta actividad incluyen intentar diferentes

formas para aprender el material, buscar explicaciones de los profesores e identificar errores de razonamiento. En el contexto de la solución de problemas la evaluación incluye un análisis de la efectividad de la ejecución y de las estrategias que puede conducir al uso de nuevas estrategias o a reformular diseños de secuencia de trabajo en sub-problemas, además da información acerca de la calidad de la solución y se puede considerar que va más allá del chequeo debido a que juzga el estado y curso del pensamiento mediante la aplicación de criterios (Davidson, Deuser y Sternberg, 1994; Kluwe, 1987).

1. 6. 2. Factores afectivos y conativos

El interesante trabajo de Snow, Corno y Jackson (1996) presenta una taxonomía muy completa con las categorías que pueden considerarse como factores que influyen en el desempeño del sujeto en una tarea.

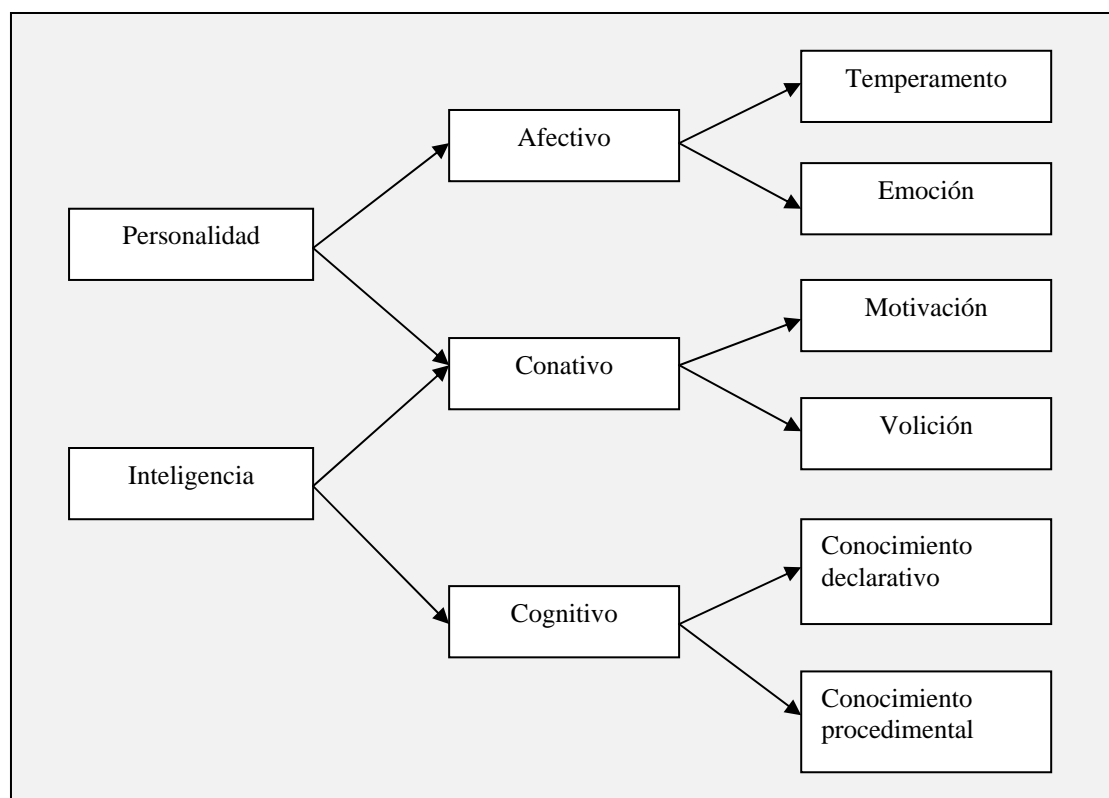


Figura 1. Taxonomía de constructos determinantes de las diferencias individuales

En esta taxonomía se consideran tres aspectos de la mente humana: lo cognitivo, lo conativo y lo afectivo, enfatizando las funciones conativas y afectivas en el área de la educación y su influencia en el aprendizaje y el rendimiento académico. Considera la inteligencia y la personalidad como los dos factores generales determinantes de las diferencias individuales en la ejecución de tareas, en la que lo cognitivo depende fundamentalmente de la inteligencia, lo afectivo de la personalidad y lo conativo está influido por esas dos categorías más generales. En la Figura 1 se ofrece una presentación esquemática de esta taxonomía.

1. 6. 2. 1. Factores afectivos

El aspecto afectivo está relacionado con los sentimientos y emociones que se producen como respuesta ante un objeto o idea particular, como la reacción general hacia algo que gusta o disgusta. Tiene dos componentes representados por el temperamento y la emoción.

Temperamento

Se refiere a un fenómeno característico del individuo que incluye características tales como susceptibilidad a la estimulación emocional, fortaleza habitual, velocidad de respuesta y humor prevalente. Usualmente se refiere a características constitucionales que son poco dependientes de las situaciones. Los constructos asociados a temperamento se relacionan con el nivel de reactividad o energía conductual y con rasgos de personalidad.

En estudios sobre temperamento se señala que personas con mayor fortaleza y velocidad de respuesta (altamente reactivas) evitan estados de estrés que reduzcan su productividad, prefieren tareas concretas, buscan estrategias auxiliares para la ejecución o invierten más esfuerzo, mientras que las de menor fortaleza y baja velocidad de

respuesta (reactividad baja) prefieren tareas abstractas, no hacen recesos y más bien buscan la novedad y la complejidad (e. g. Farley, 1985; Strelau, 1983).

Emoción

La emoción es un estado de ánimo producido por elementos situacionales que se puede traducir en actitudes o conductas. Aunque la emoción usualmente se relaciona con estados, no con rasgos, pueden llegar a producirse tendencias en determinadas situaciones. El humor emocional característico es un constructo asociado a esta categoría afectiva. Watson, Clark y Tellegen (1988) clasifican el humor emocional característico en humor positivo, medio y negativo. Con humor positivo tendremos un sujeto alerta, atento, que pone empeño, interesado y orgulloso; con humor negativo un sujeto es temeroso, vergonzoso, hostil y nervioso y en la categoría media un sujeto cauteloso, reflexivo, laborioso y sensible. Los estudios que investigan las relaciones entre humor emocional característico y funciones cognitivas (e. g. Bower, 1981; Isen, Daubman y Gorgoglione, 1987; Pekrun, 1994) muestran como los estudiantes que difieren en su humor característico difieren cualitativamente en lo que aprenden; por ejemplo, los estudiantes con humor característico positivo codifican la información nueva de modo más creativo, tienen más facilidad para el aprendizaje y para desempeñarse con éxito en situaciones de solución de problemas.

Cabe también señalar que algunos trabajos sobre factores de personalidad reseñados por Snow y cols. hacen especial énfasis en extraversión vs. introversión y neuroticismo como elementos que marcan el temperamento, así como también en la impulsividad y la ansiedad como variables subordinadas a los factores de personalidad que influyen en el aspecto emocional y que pueden afectar de manera importante los procesos de solución de problemas y la toma de decisiones.

1. 6. 2. 2. Factores conativos

El aspecto conativo está relacionado con las conductas de toma, mantenimiento y ejecución de decisiones y cursos de acción para el logro de metas. La motivación y la volición representan el aspecto conativo del funcionamiento psicológico humano. Los procesos motivacionales median la toma de una decisión y la promueven y los volitivos median la puesta en marcha de esas decisiones.

Motivación

El dominio motivación está relacionado con la toma de decisiones y la selección de metas individuales. Puede considerarse el estado predecisional (Heckhausen, 1991). La motivación se hace presente en una gran cantidad de modelos sobre rendimiento, habilidad, desarrollo de la experiencia y solución de problemas (véase por ejemplo Kupermintz, 2002; Lau y Roeser, 2002; Nickerson, 1994; Snow, 1989; Sternberg, 1998a y Sugrue, 1995). La motivación se refiere a deseos, apetencias, necesidades y metas, de modo que se pueden situar en esta categoría variables como orientación hacia el logro, auto-concepto, auto-eficacia, valor de la tarea, expectativas e intereses.

El trabajo de Pintrich y Schunk (2002) sobre la motivación como factor determinante del aprendizaje y de la ejecución resalta la importancia de algunas teorías desde las cuales se ha aproximado el estudio de este factor, en particular el modelo expectativa-valor de Eccles y Wigfield (1995) y la teoría cognitivo social de Bandura (1997).

El modelo expectativa-valor (Eccles, 1983; Wigfield, 1994; Eccles y Wigfield, 1995, 2002) se centra en el rol de las expectativas de los estudiantes sobre su éxito académico y el valor percibido de la tarea como los dos más importantes predictores de conductas de logro. El constructo expectativa se refiere a la respuesta del examinado a la pregunta ¿soy capaz de hacer esta tarea?, que involucra juicios sobre su competencia

y acerca de la dificultad percibida de la tarea. El constructo valor percibido de la tarea se refiere a la respuesta del estudiante a la pregunta ¿por qué debería hacer esta tarea?, la cual puede incluir interés en el tópico, creencias sobre su utilidad e importancia y beneficios que pueda reportarle la ejecución de la misma.

En la teoría cognitiva social de Bandura (1993, 1997) la auto-eficacia y las expectativas sobre los resultados son centrales como determinantes de la conducta. Define la auto-eficacia como el juicio del individuo sobre sus propias capacidades para organizar y ejecutar cursos de acción requeridos para el logro de metas fijadas, lo que permite asociarla con el conocimiento autopercebido señalado dentro de la categoría metacognitiva. La auto-eficacia no se refiere a las destrezas que el sujeto tiene sino a lo que cree que puede hacer con lo que tiene en una diversidad de circunstancias, operando como un factor clave en los procesos motivacionales.

Volición

El dominio volición está relacionado con los constructos involucrados en la consecución de metas y con las acciones seguidas para llevar a cabo, de la mejor manera, planes e intenciones. El estado postdecisional será lo que se considere como volición puesto que se da una vez que la acción comienza (Heckhausen, 1991). La volición se refiere a intenciones, esfuerzo, acciones y auto-regulación, de modo que se pueden situar aquí las variables orientación y control de acciones, aprendizaje auto-regulado y estilos de aprendizaje (Jackson, 1994; Kuhl y Beckmann, 1985; Kupermintz, 2002; Schunk y Zimmerman, 1994; Willingham, 1985).

La manera de orientar las acciones puede ser orientación acción vs. orientación estado u orientación dominio vs. orientación ejecución. Según la teoría de Kuhl (1985), un individuo orientado hacia la acción tiende a actuar en lo inmediato para llevar a cabo sus propósitos, mientras que el individuo orientado hacia el estado mide sus acciones y

se ve afectado notablemente si en el pasado ha fallado. Las personas orientadas hacia el dominio buscan retos y mantienen el control aun en caso de fallar, se esfuerzan para aumentar sus competencias, mientras que las personas orientadas a la ejecución evitan los retos, prefieren tareas conocidas y se pueden ver afectadas en situaciones de fallo (Dweck y Leggett, 1988). Kuhl (1985) establece que el control de las acciones es clave para manejar competentemente los distractores que pueden afectar los procesos atencionales que permitirán al sujeto dedicarse a la tarea y emplear los recursos disponibles oportuna y eficientemente.

El aprendizaje auto-regulado se observa cuando los sujetos son capaces de aplicar deliberadamente estrategias que promueven un aprendizaje más profundo (Kuhl, 1984). Las estrategias utilizadas son las mencionadas dentro del apartado relativo a metacognición, en particular, los procesos de control metacognitivo, que son considerados por este autor dentro de lo volitivo al ser acciones que contribuyen al logro de la meta.

El estilo personal de aprendizaje es considerado una categoría volitiva en cuanto que constituye una fuente de diferencias individuales en aprendizaje y solución de problemas. Se puede definir el estilo personal como una estrategia que se usa constantemente para un tipo de tarea, o como la forma habitual o preferida que adopta un individuo para encarar situaciones sociales o cognitivas. En este sentido, los estilos de aprendizaje involucran un enfoque para aprender y estudiar que puede ser profundo o superficial, orientado a lo reflexivo o a lo operativo y que puede influir en las conductas implicadas en la ejecución de la tarea (Snow et al., 1996).

En este sentido Kyllonen, Walters y Kaufman (2005) señalan que los estudios sobre el test *Graduate Record Examination* (GRE), una prueba que se aplica a las personas que se postulan a estudios de postgrado en los Estados Unidos, han llevado a

la conclusión de que la validez de esta prueba podría ser mejorada incluyendo variables no cognitivas tales como personalidad e intereses. Estos autores revisan los factores de personalidad estableciendo que un sujeto neurótico tiene más posibilidades de abandonar la tarea que otro, o que un sujeto con apertura a nuevas experiencias tiene motivaciones distintas en función del tipo de tarea que se le presente, todo lo cual puede afectar el éxito en graduados y su rol potencial en la admisión para el empleo.

1. 6. 3. Factores relativos a la tarea

Todo lo relacionado con la forma de aplicación, las instrucciones y organización del test, el formato y orden de las preguntas, el hecho de que el test sea de velocidad o de potencia, puede considerarse como factores que influyen en alguna medida en la ejecución. Actualmente los tests pueden ser aplicados en la forma tradicional de lápiz y papel o en forma computarizada, ya sea vía internet o que simplemente deba responderse en el ordenador. Las instrucciones pueden ser señaladas en forma oral o aparecer en el test. Cualquiera de estas características es susceptible de producir diferencias en la ejecución de la tarea que lleva a cabo el examinado.

Pero particularmente el ítem, su contenido y su forma, es el que debe elicitar la conducta que representa el constructo que busca ser medido. Son numerosos los estudios que se han realizado sobre la posible influencia del formato del ítem y numerosas las propuestas de seguir investigando no solo en lo que respecta al análisis cognitivo, sino también en lo que tienen que ver con aspectos motivacionales y afectivos que podrían ser mediados por la forma de presentación de la tarea (Martínez, 1999; Snow, 1993).

1. 7. EL FORMATO DEL ÍTEM

El ítem, tal y como lo define Osterlind (1998), es una unidad de medida con un estímulo y una forma preceptiva de responder para producir una respuesta del examinado, a partir de la cual puede ser estimado su nivel en el constructo psicológico medido. El formato del ítem, por su lado, refiere a su diseño y disposición, lo cual hace que su categorización pueda resultar muy diversa.

En términos generales, se suelen considerar dos categorías amplias de formatos de ítems: (a) ítems de respuesta seleccionada y (b) ítems de respuesta construida (Bennett y Ward, 1993; Downing, 2006; Osterlind, 1998). Los ítems de respuesta seleccionada generalmente están compuestos por un enunciado que representa un problema, plantea una pregunta o recoge una afirmación, seguida por un conjunto de alternativas de respuesta, en donde una o varias representan la respuesta correcta y el examinado debe seleccionar la respuesta que considere acertada. En contraste, los ítems de respuesta construida no suministran ninguna respuesta, sino que el examinado debe elaborar la respuesta a través de la interpretación, análisis y utilización de la información suministrada por el ítem y de su conocimiento y experiencia previa.

Los ítems de elección múltiple y los de verdadero y falso son ejemplos típicos del formato de respuesta seleccionada, siendo el primero ampliamente utilizado en las pruebas de evaluación educativa a gran escala por las ventajas que proporciona. Se señalan entre ellas su bajo costo operativo, su facilidad de calificar objetivamente, su mayor precisión y la posibilidad de contar con muestras representativas del dominio evaluado (Mumford, Baughman, Supinski y Anderson, 1998; Osterlind, 1998; Ryan y Greguras, 1998). No obstante, en las últimas décadas este formato pierde exclusividad en las pruebas utilizadas en macroencuestas educativas, encontrándose algunas

debilidades en el mismo. Por un lado, se afirma que es limitado para evaluar habilidades de nivel cognitivo superior, para evaluar destrezas para el aprendizaje a lo largo de la vida y para poder llevar a cabo la más que conveniente integración de la evaluación con la instrucción, de acuerdo con las teorías del aprendizaje y la psicología cognitiva que se manejan en la actualidad. Por otro lado, se señala que para medir destrezas psicomotoras puede resultar inútil y que para tareas complejas como ejecución de instrumentos o destrezas en disciplinas artísticas solamente el formato de respuesta construida es el adecuado (Downing, 2006; Martínez Arias, 2010). Además, las respuestas en este formato son muy fáciles de ser copiadas fraudulentamente o de ser respondidas al azar.

El formato de respuesta construida adopta una gran variedad de formas, tales como completar oraciones, reordenar frases o secuencias, responder con una palabra o un número, emitir una respuesta argumentada, resolver problemas, hacer experimentos o hacer muestras de escritura. Los ítems con este formato pueden, a su vez, ser clasificados en dos grandes categorías representadas por los ítems de respuesta corta y por los ítems de respuesta extendida. Los ítems de respuesta corta sólo exigen como respuesta una palabra, una frase o un número, mientras que los ítems de respuesta extendida requieren ejecuciones más complejas tales como escritura de ensayos, solución de problemas, realización de experimentos científicos u otras modalidades de lo que se conoce con el término anglosajón de *performance assessment* o evaluación de la actuación. Al igual que los ítems de respuesta seleccionada los ítems de respuesta construida tienen virtudes y debilidades.

Aun cuando se dice que la demanda en cuanto a procesos cognitivos involucrados para llegar a la respuesta es superior que la que implica la selección de una respuesta entre varias, los ítems de respuesta corta adolecen de muchas de las críticas

realizadas a los ítems de elección múltiple, por cuanto también se les considera diferentes de lo que podría ser un problema de la “vida real”.

Por su parte, los ítems de respuesta extendida resultan bastante atractivos, no sólo porque parecen capturar habilidades complejas, sino también porque permiten la observación de los procedimientos llevados a cabo por un sujeto para emitir una respuesta, lo cual posibilita establecer diferencias cualitativas entre los evaluados y ayuda a obtener información para fines de diagnóstico. Particularmente los ítems tipo evaluación de la actuación permiten la alineación de destrezas y habilidades con las competencias importantes para la vida en contextos realistas (Martínez Arias, 2010). No obstante, este formato también tiene sus detractores, que basan sus argumentos en deficiencias tales como que la escala para calificar las pruebas o productos plantea problemas de objetividad y que la mayor demanda de tiempo no permite cubrir adecuadamente el dominio a evaluar, repercutiendo en la validez y limitando la generalizabilidad de las inferencias (Bennett, 1993; Martínez, 1999; Messick, 1998; Mumford et al, 1998). Además, su aplicación y calificación resulta muy costosa y tiende a centrarse más en el producto o proceso que en el constructo de interés

Los estándares para la evaluación psicológica y educativa (AERA, APA y NCME, 1999) indican que el formato del ítem debe especificarse en función del propósito del test. Por tanto, es necesario conocer a fondo cómo funciona cada uno de los formatos que puedan ser empleados en los procesos de medida de los constructos psicológicos y educativos.

Existen diversas publicaciones que tratan acerca de las características, naturaleza y potencialidades de los distintos formatos del ítem (Downing y Haladyna, 2006; Haladyna, 1997, 1999; Kane y Mitchell, 1996; Osterlind, 1998), así como también estudios acerca de la equivalencia del constructo cuando es medido con diferentes

formatos (Ackerman y Smith, 1988; Ayala, Shavelson, Yin y Shultz, 2002; Bennett, Rock y Wang, 1991; Bridgeman, 1992; Jodoin, 2003, entre otros).

Ackerman y Smith (1988) establecieron la estructura factorial teórica del constructo a medir y luego compararon los resultados de los análisis factoriales confirmatorios, a fin de observar el grado de correlación entre cada factor en función del formato del ítem, encontrando que al menos en escritura el constructo medido es función del formato del ítem, puesto que los puntajes obtenidos con los distintos métodos (ensayo, respuesta abierta y elección múltiple) proporcionan diferente información cognitiva y concluyen que una medida más confiable de este constructo puede requerir el uso de varios formatos.

Ayala, Shavelson, Yin y Shultz (2002) realizaron un estudio con el propósito de determinar si pruebas de evaluación de la actuación podían ser explícitamente diseñadas para medir tres dimensiones del razonamiento en la ciencia (conocimiento básico, razonamiento mecánico espacial y razonamiento cuantitativo) y, además, examinar la consistencia entre tres medidas del rendimiento, obtenidas mediante ítems de elección múltiple, respuesta abierta y evaluación de la actuación. Encontraron una fiabilidad muy baja en los ítems de respuesta abierta y por lo tanto no los consideraron para el resto de sus análisis. También hallaron que los puntajes obtenidos con los ítems de elección múltiple y de evaluación de la actuación no convergen en las tres dimensiones del razonamiento hipotetizadas, obteniendo además correlaciones muy moderadas que sugieren que podrían medir aspectos diferentes del rasgo considerado.

Bennett, Rock y Wang (1991) evaluaron la equivalencia del constructo medido con ítems de elección múltiple y de respuesta abierta en el área de computación, a través de un modelo de dos factores correlacionados (el primer factor para los ítems de elección múltiple y el segundo factor para los ítems de respuesta abierta), encontrando

que un único factor provee el ajuste más parsimonioso. Sin embargo, los autores advierten que la evidencia encontrada es limitada y los resultados no pueden ser generalizados.

Bridgeman (1992) comparó distintos formatos de respuesta en lo concerniente a dificultad, discriminación y estructura correlacional utilizando la TRI y encontró resultados contradictorios. Por ejemplo, aunque a nivel de los ítems se encontraba mucha diferencia en cuanto a dificultad, las estimaciones de los parámetros de los sujetos con los distintos formatos eran muy similares; por otro lado, las curvas características de los ítems en algunos casos se superponían y en otros eran muy distintas.

Jodoin (2003) realizó una investigación con el objetivo de proveer evidencia empírica de la fiabilidad de dos tipos de ítem, el clásico de elección múltiple y un tipo nuevo de respuesta construida que se deseaba probar en el examen computarizado del *Microsoft Certified Systems Engineer*. La confiabilidad fue evaluada usando las funciones de información de los ítems y se pudo concluir que los ítems de formato novedoso proveen considerablemente más información que los ítems de elección múltiple para todos los niveles de habilidad. No obstante, responder estos ítems lleva más tiempo y en consecuencia proveen menos información por minuto que los de elección múltiple, lo que puede tener mucha importancia al momento de seleccionar el tipo de ítem más adecuado para una determinada evaluación.

En un estudio sobre pruebas de acceso a la Universidad, Sternberg y colaboradores (2004) encontraron que los tests en formato de elección múltiple conformaron un factor, independientemente del objetivo de los mismos, lo que los llevó a plantear la importancia del estudio del formato. Los autores consideran que si el constructo medido con diferentes tipos de formatos no es equivalente, se debe analizar

si el formato representa una fuente de varianza irrelevante al constructo (varianza método) o si verdaderamente los ítems elaborados en formatos diferentes están midiendo cosas distintas representativas del mismo constructo.

También pueden hallarse investigaciones como las de Hogan (1981), Traub (1993), Martínez (1999) o Rodríguez (2002, 2003), que recogen los resultados de variados estudios empíricos sobre formatos de ítem. En todos estos trabajos se señala la importancia del estudio del formato en la determinación de las diversas fuentes de variación, la existencia de diferencias y, muy importante, que la elección del formato del ítem está estrechamente vinculado a la representación del constructo a medir, en consecuencia, es un asunto crucial para garantizar la validez de las inferencias.

Hogan (1981) realiza una amplísima revisión de investigaciones sobre estas cuestiones y concluye que los ítems planteados en formato de elección miden lo mismo que los formulados en formato de construcción. Sin embargo, acota que no se deben hacer generalizaciones debido a que la literatura seleccionada para la revisión está limitada al campo de la medición de lo cognitivo, fundamentalmente conocimientos en dominios académicos distintos a lectura y escritura. También matiza sus resultados advirtiendo acerca de la poca diversidad de la población estudiada y establece otras limitaciones de su estudio relacionadas con los métodos que fueron utilizados para evaluar la equivalencia.

Traub (1993) selecciona nueve estudios sobre la equivalencia del rasgo medido, considerados por el autor los trabajos mejor diseñados y conducidos de las publicaciones más recientes para la fecha de su investigación. El propósito del estudio fue identificar los hallazgos consistentes y recomendar futuras líneas de investigación. La evidencia encontrada no es suficientemente sólida para afirmar que distintos formatos puedan medir lo mismo -en algunos dominios se revela equivalencia, mientras

que en otros no- y tampoco hay hallazgos que indiquen dónde reside la diferencia. Recomienda para las próximas generaciones de estudios no considerar específicamente el asunto de la equivalencia sino más bien el del efecto del formato.

En su investigación, ampliamente citado en la literatura sobre el tema, Martínez (1999) trabaja sobre la hipótesis de que los ítems de selección y de respuesta construida difieren no sólo en la demanda cognitiva sino también en el rango de cognición a que cada uno de ellos puede dar lugar. Basándose en los resultados de trabajos empíricos previos señala la estrecha relación entre el formato del ítem y lo que se mide con éste, concluyendo que no existe un formato apropiado para todo propósito en toda ocasión. Afirma que usar una combinación de formatos puede minimizar la varianza método atribuible a un formato específico.

Rodríguez (2002, 2003) realiza un interesante trabajo de revisión meta-analítica que recoge más de 60 investigaciones sobre el formato del ítem realizadas en alrededor de los últimos 80 años, basando la importancia de este tipo de investigaciones en tres razones: (1) las interpretaciones varían de acuerdo al formato del ítem, (2) las diferencias en el costo son importantes y (3) las consecuencias de usar un formato dado pueden afectar la instrucción. En la revisión de los trabajos, que emplean la correlación u otras técnicas para el estudio de la equivalencia del formato, encuentra valores altamente heterogéneos que no permiten ser concluyente. Partiendo de estos resultados hace algunas consideraciones en relación a la selección del formato, relacionadas con el costo y con aspectos políticos, para finalmente concluir que lo fundamental es diseñar los tests de forma tal que los ítems puedan dar lugar a la clase de conductas especificadas en la definición del constructo en una vía en la cual la validez quede garantizada, subrayando de esta forma la importancia de la investigación acerca del significado de las puntuaciones aportadas por un formato en particular de ítem.

Otro material fundamental en este tema es el libro editado por Bennett y Ward (1993), en el que se recoge un conjunto de 14 trabajos cardinales sobre el formato de los ítems. Messick (1993), por ejemplo, aconseja realizar estudios que posibiliten distinguir entre ítems que en función de su formato puedan presentar diferencias en relación a demandas de procesamiento de información. Snow (1993) propone un conjunto de 8 hipótesis rivales en las que plantea que las diferencias en los formatos pueden o no estar en la actitud hacia los mismos, en la ansiedad que pueden producir al examinado durante la ejecución, en la motivación que generan, en la forma como influyen la instrucción y el aprendizaje, en las habilidades y estructuras de conocimiento que son capaces de medir, en la profundidad del procesamiento cognitivo y en la adecuación psicométrica de cada uno a las condiciones de prueba. Bennett (1993) explora el término respuesta construida atendiendo a los significados que pueda tener desde el punto de vista descriptivo, de su validez y su fiabilidad. En fin, este libro provee una útil compilación de referencias sobre el tema y ofrece un marco para el desarrollo de la investigación sobre el formato de los ítems como un elemento de gran importancia en la evaluación cognitiva.

La revisión de estas publicaciones conduce a pensar que la investigación en este campo todavía tiene mucho que aportar. Pareciera que los ítems de selección no miden exactamente lo mismo, ni con la misma precisión que los ítems de respuesta construida; que la naturaleza de las diferencias entre unos y otros no ha sido todavía comprendida en toda su extensión; que la profundidad cognitiva que puede evocar cada tipo de formato difiere; que los procesos metacognitivos que activan son diferentes y que las variables afectivas y conativas funcionan diferencialmente en función del formato. Se percibe de lo anterior un llamado a continuar profundizando en la investigación acerca del formato del ítem, por el papel tan importante que juega en la investigación sobre

validez. Conociendo la naturaleza de lo que logra medir cada formato será más sencillo determinar cuál será el más adecuado para medir el constructo de interés con un determinado objetivo y para un grupo de sujetos y contexto de aplicación particulares.

CAPÍTULO II

UN MODELO TEÓRICO PARA SOMETER A PRUEBA

No hay ciencia si no se inventa, no hay ciencia si no se justifica.

Mariano Yela

La psicología cognitiva plantea que el saber es más que la mera acumulación de conocimiento factual y la utilización de procedimientos rutinarios: significa ser capaz de integrar conocimientos, destrezas y procedimientos de forma que sean útiles para interpretar situaciones y resolver problemas (Pellegrino, Chudowsky y Glaser, 2001). Por ello, se considera la ejecución de tareas mentales a dos niveles: cómo se procesa la información y cómo se organiza el conocimiento específico asociado a dominios particulares del conocimiento.

Los constructos asociados a la solución de problemas (aptitud, inteligencia, rendimiento) no son unidimensionales sino que reflejan una compleja combinación de procesos, estrategias, estructuras del conocimiento y otras variables no cognitivas. Son muchas las fuentes que determinan la dificultad de una tarea y deben ser especificadas tal que puedan ser sistemáticamente manipuladas para investigar los distintos procesos cognitivos implicados. Como se ha puesto claramente de manifiesto en las investigaciones examinadas anteriormente, las diferencias individuales no están sólo en las aptitudes, en cada componente del procesamiento de información o en los conocimientos previos, sino también en el uso de estrategias al resolver la tarea en cuestión, en los recursos psicológicos básicos de que disponga cada sujeto, en sus motivaciones, en la manera de percibirse a sí mismos y en la forma como pueden ser influidos por variables contextuales relacionadas con la tarea.

El objetivo de la presente investigación doctoral es aproximar una explicación acerca del significado psicológico de las puntuaciones obtenidas cuando el constructo es medido con formatos diferentes: lo que se pretende es investigar donde reside la diferencia (si la hubiere) entre los distintos formatos, para lo cual es necesario examinar esas variables de distinta naturaleza que explican la actuación de los sujetos en el ítem y

que pueden dar cuenta de las diferencias individuales que subyacen en la puntuación obtenida por el sujeto.

Para este examen se partió de un modelo teórico cuyo planteamiento es claramente deudor de las propuestas de Snow, Corno y Jackson (1996) y que además incorpora aportes de otros autores que van en la misma línea (e.g. Benton y Kiewra, 1987; Masters y Mislevy, 1993; Nickerson, 1994; O'Neil y Schacter, 1997; Schoenfeld, 1983; Sternberg, 1985, 1998b), al considerar la importancia de variables de diversa índole en la resolución efectiva de problemas. Con este modelo se pretende explorar cómo resuelve el sujeto tareas planteados en distintos formatos, en función de la gestión que hace de sus habilidades cognitivas y no cognitivas.

En el modelo propuesto se incorpora un conjunto importante de las variables que las distintas investigaciones precedentes señalan como los factores influyentes más significativos, buscando un modelo parsimonioso, que permita manejar adecuadamente la información disponible, pero que resulte lo más completo e informativo posible.

En la Figura 2 se presenta un esquema del modelo propuesto. Como se puede observar, las variables provenientes de la dimensión cognitiva que se incluyen son el conocimiento (declarativo y de procedimientos) y los componentes de procesamiento de la ejecución cognitiva. Las variables incluidas que pertenecen al ámbito no cognitivo son (1) las referentes a los aspectos metacognitivos (en particular, el conocimiento estratégico, la regulación metacognitiva y la percepción del sujeto sobre sí mismo) y (2) las relativas a la tarea, tanto desde la perspectiva del sujeto (en cuanto al valor o dificultad percibida en la misma), como desde lo relacionado directamente con el formato en que se presente la tarea.

Lo que supone este planteamiento es que los procesos mentales –cognitivos y no cognitivos- que se activan en el sujeto cuando se enfrenta a una tarea se pueden

manifestar diferencialmente en función del formato del ítem.

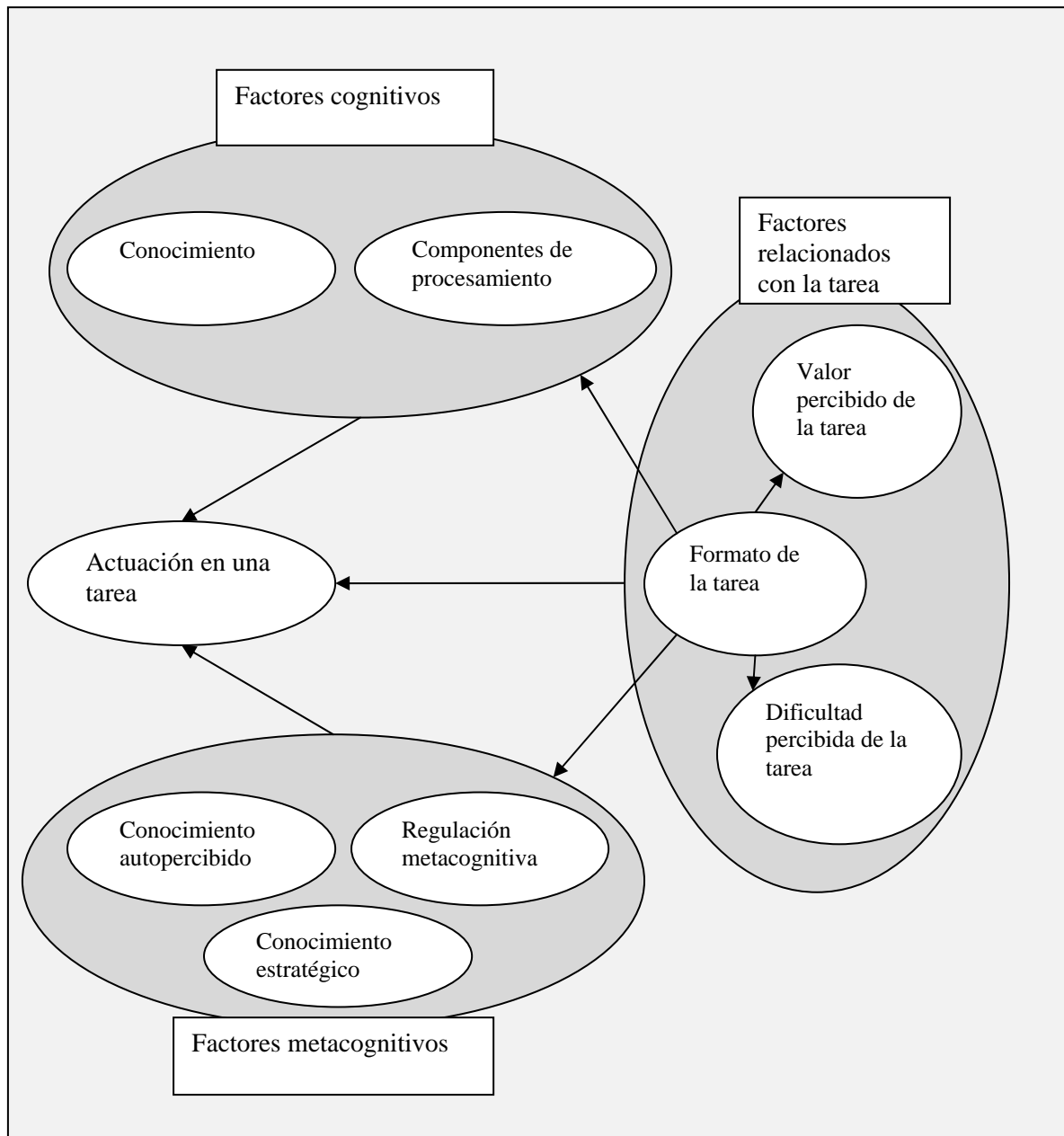


Figura 2. Modelo propuesto

Cuando el sujeto responde al ítem, utiliza sus conocimientos acerca del dominio específico al que se enfrenta y los procedimientos que ayuden a la solución del problema, recurriendo a las estrategias o heurísticos que considere adecuadas para cada

situación, así como poniendo en juego su habilidad para controlar todas las actividades cognitivas implicadas en la solución del problema en cuestión, planificando su forma de trabajo, controlando la efectividad de su plan y verificando la plausibilidad de su respuesta. Estas acciones se ven influidas por sus creencias acerca de su competencia para resolver la tarea, por la importancia o valor atribuida a ésta y por su percepción acerca de la dificultad de la misma, todos ellos factores decisivos en la motivación del sujeto y, en consecuencia, en su éxito para resolverla; con la posibilidad de que se revelen diferencias individuales dependiendo del formato de la tarea.

2. 1. FACTORES COGNITIVOS

Dentro de este ámbito, la selección de las variables a incluir en el modelo se realizó en consonancia con el planteamiento fundamental de la psicología cognitiva, que propone como principales elementos en la actuación la organización del conocimiento y el procesamiento de la información.

El estudio de las estructuras del conocimiento tiene una relevancia fundamental y representa el más importante aporte de las teorías cognitivas a la medición en el ámbito educativo, al contribuir a la descripción de la adquisición, organización y uso del conocimiento en dominios específicos (Pellegrino, Chudowsky y Glaser, 2001). Particularmente el conocimiento declarativo y el conocimiento procedimental son considerados esenciales para una ejecución exitosa, constituyéndose en categorías de la arquitectura del conocimiento ampliamente aceptadas, las cuales siempre son incluidas como las formas fundamentales de organización y almacenamiento del conocimiento. Además, se incorpora una medida del conocimiento en matemáticas, considerado una variable importante para la resolución de problemas.

Por su parte, el estudio de los componentes del procesamiento cognitivo es absolutamente clave en las investigaciones cognitivas, abordándose desde diversos enfoques. En particular, nos parece especialmente apropiado el enfoque propuesto por Sternberg (1985, 1998b) en su teoría triárquica de la inteligencia, que considera la existencia de un numeroso conjunto de procesos que se activarán en función de la tarea que deba realizarse y que se agrupan, de modo resumido, en tres fases básicas: la codificación (el sujeto representa mentalmente la tarea identificando los elementos que permitirán la solución del problema), la combinación (el sujeto realizará las operaciones necesarias para alcanzar y valorar una solución) y respuesta (se expresa el resultado de las fases anteriores).

2. 2. FACTORES METACOGNITIVOS

La relevancia de la metacognición como un determinante en la ejecución de tareas la posiciona como un factor que debe ser incluido dentro del modelo que se propone, pues estos procesos actúan como guía en el proceso y mejoran la eficiencia de las conductas orientadas hacia la solución de problemas (Davidson, Deuser y Sternberg, 1994).

La planificación y el control ejercido durante la ejecución de la tarea son importantes elementos de regulación cognitiva que contribuyen a un mejor desempeño y que son ampliamente reconocidos en numerosos estudios (e. g. Davidson, Deuser y Sternberg, 1994; Justice y Dornan, 2001; Nickerson, 1994; Schraw y Dennison, 1994). En su teoría triárquica Sternberg (1985, 1998b) señala que los componentes de ejecución sólo se desencadenan si son mediados por los procesos de control ejecutivo.

Por su parte, el conocimiento estratégico, representado por los heurísticos

utilizados para la solución de problemas, resulta clave para obtener la solución de modo más efectivo (Krathwohl, 2002; Nickerson, 1994; Schoenfeld, 1985).

El conocimiento autopercebido o creencias acerca de las propias competencias es una variable que se considera un determinante de gran importancia en la solución de problemas, tanto desde la perspectiva de la metacognición, como desde la perspectiva de la motivación y la volición, generando acciones cognitivas decisivas en la solución de problemas (Kupermintz, 2002; Lau y Roeser, 2002; Schoenfeld, 1983, 1985; Sugrue, 1995).

2. 3. FACTORES RELACIONADOS CON LA TAREA

El modo en que las diferencias conativas entre los estudiantes influyen la actividad cognitiva, tanto en la escuela como fuera de ella, hace de estas variables un importante elemento a explorar para poder interpretar con mayor exactitud el significado de las puntuaciones.

El repertorio de variables de este tipo que pueden influir en la conducta del sujeto ante una tarea es realmente enorme. Sin embargo, optamos por el planteamiento de Eccles y Wigfield (1995, 2002) al incluir el valor percibido de la tarea y la dificultad percibida de la tarea, pues según exponen estos autores, éstas son las variables fundamentales que promueven los procesos motivacionales y volitivos, suscitando conductas de logro o de disuasión según se considere la tarea en relación a estas variables.

El hecho de que una tarea resulte entretenida, atractiva, permita demostrar las capacidades y sea considerada de utilidad, es decir, el valor que se le dé a la tarea, contribuirá a motivar al sujeto a realizarla. Por su parte, si el sujeto percibe que el

esfuerzo y la dificultad de la tarea puedan sobrepasar sus capacidades, la voluntad de enfrentar la tarea resulta mermada. En este sentido, podemos considerar estas variables como atributos motivacionales y volitivos con influencia determinante en la ejecución del ítem.

Por su parte, el formato del ítem puede representar un factor decisivo de la actuación del sujeto frente a una tarea particular, al constituirse en elemento que puede estar asociado a diferencias individuales en las variables tanto cognitivas como no cognitivas que se activan en el momento de la ejecución (Martinez, 1999; Snow, 1989). En tal sentido, el estudio del formato refiere al asunto de la validez por cuanto existe la preocupación de examinar si distintos formatos son capaces de medir un constructo al mismo nivel de amplitud y profundidad. Tal como afirma Messick (1993), la equivalencia del rasgo medido con distintos formatos es un tema concerniente con la validez de constructo, de tal modo que las investigaciones sobre el formato del ítem y su influencia en la ejecución de la tarea se perfilan como un importante aporte en el estudio del significado de las puntuaciones.

CAPÍTULO III

MÉTODO

Es más fácil morir con honra que pensar con orden.

José Martí

3. 1. INSTRUMENTOS

Para la comprobación empírica del modelo propuesto se procedió a la preparación de los instrumentos necesarios para la recolección de los datos. Se realizó una búsqueda exhaustiva de instrumentos diseñados para medir los constructos que componen el modelo a someter a prueba en esta investigación. Aunque no se utilizó ninguno de los instrumentos hallados, la revisión de los mismos resultó de gran utilidad para la elaboración de las pruebas necesarias para medir las variables implicadas en el modelo.

Los instrumentos utilizados en esta investigación son los siguientes:

1. La prueba de competencias, para observar la ejecución de los estudiantes en cada tipo de formato de ítem.
2. La prueba general, para obtener información sobre un conjunto importante de variables del modelo.
3. La hoja de registro de los profesores, para estimar el nivel de conocimiento de los estudiantes.

3. 1. 1. Prueba de competencias

La prueba de competencias permitiría obtener la respuesta del sujeto a cada uno de los ítems formulados en tres distintos tipos de formato. Se optó por incluir un solo ítem por cada tipo de formato considerado en el estudio con el objeto de optimizar el tiempo de aplicación y de esta manera conseguir la máxima colaboración de los responsables de los centros escolares, al interrumpir lo menos posible el desarrollo de la dinámica escolar. Los formatos utilizados en esta prueba son: elección múltiple (EM), respuesta corta (RC) y evaluación de la actuación (EA).

La elección de las preguntas de la prueba de competencias fue una tarea que implicó revisar numerosos ítems que, además de estar formulados en los distintos formatos de interés, tuvieran como característica imprescindible el no estar estrechamente vinculados a un determinado programa de estudios y que cubrieran condiciones impuestas propiamente por la investigación, como similitud en los procesos cognitivos implicados en la ejecución de la tarea planteada, bajo costo y mínimo tiempo de aplicación.

Dado que el estudio se centra en el formato y no en el contenido del ítem, el área específica de conocimientos a elegir no revestía un carácter fundamental: los ítems debían tratar sobre contenidos conocidos para los estudiantes, cuyo conocimiento hubiese sido adquirido –en mayor o menor grado- a lo largo del proceso de formación académica del estudiante y con los que éste estuviera familiarizado en el momento de la evaluación.

Al presentar al sujeto un solo ítem por formato se necesitaba que, en lo posible, comportaran el uso de los mismos componentes de procesamiento cognitiva y de esta manera hacer más plausibles las comparaciones.

Por otra parte, el hecho de necesitar un tamaño muestral suficiente para garantizar la posibilidad de aplicar determinadas técnicas estadísticas significaba una limitación en cuanto a la elección del ítem de evaluación de la actuación, puesto que si este requería el empleo de materiales u otros recursos de los que el estudiante no dispone habría sido necesario suministrárselos, con el consiguiente encarecimiento de la actividad y la necesidad de emplear tiempo y esfuerzos en la consecución de esos recursos.

Las razones que justifican orientar la búsqueda de ítems hacia aquellos cuya ejecución no acarreará mucho tiempo son, además de la indicada anteriormente,

producir la menor fatiga en el estudiante para que pudiera completar los tests a satisfacción.

En este proceso se consideraron tanto los ítems utilizados en las pruebas del *Third International Mathematics and Science Study* (TIMSS) como los de las pruebas para el *Programme for International Student Assessment* (PISA), dos estudios comparativos internacionales con gran solvencia dentro de la comunidad científica. Los ítems utilizados en la edición 2003 del estudio PISA (INECSE, 2005) resultaron ser más adecuados para los propósitos de la investigación, dado que estos reunían en mayor medida las características antes señaladas, particularmente los del área de conocimiento solución de problemas, al no estar asociados directamente a ningún área curricular concreta y no demandar recursos especiales que encarezcan los costos de aplicación ni tiempos excepcionales para realizar la tarea. Así, la solución de problemas resulta ser el dominio evaluado en la prueba de competencias.

Se procedió, en primer lugar, a una preselección de ítems candidatos y, en segundo lugar, a una revisión detallada de estos ítems para llegar a la elección definitiva de un ítem por formato. Para representar el ítem de evaluación de la actuación se utilizó la pregunta 8 del problema denominado *Campamento* de la prueba correspondiente al dominio de solución de problemas (véase la Figura 3). Para los dos restantes ítems se seleccionó el problema identificado como *Vacaciones* introduciendo unas mínimas modificaciones en la tabla de distancias, para forzar al estudiante a realizar algunos cálculos y poder utilizar el mismo enunciado para la pregunta de elección múltiple. En particular, se utilizó la pregunta 15 de la prueba sobre solución de problemas como ítem de respuesta corta (véase Figura 4) y se elaboró el ítem de elección múltiple que se presenta en la Figura 5. Para asignar puntuaciones a cada ítem se procedió tal como se hace en el estudio PISA.

El Departamento de Servicios Sociales de Zedlandia está organizando un campamento de cinco días para jóvenes. Se han inscrito en el campamento 46 jóvenes (26 chicas y 20 chicos) y 8 adultos voluntarios (4 hombres y 4 mujeres) atenderán y organizarán el campamento.

ADULTOS
Beatriz
Carolina
Olga
Patricia
Esteban
Ricardo
Guillermo
Pedro

HABITACIONES	
Nombre	Número de camas
Roja	12
Azul	8
Verde	8
Púrpura	8
Naranja	8
Amarilla	6
Blanca	6

Normas de las habitaciones:

1. Chicos y chicas deben dormir en habitaciones separadas
2. Al menos un adulto debe dormir en cada una de las habitaciones
3. El adulto que duerma en cada habitación debe ser del mismo sexo que el de los jóvenes

Complete la tabla colocando a los 46 jóvenes y a los 8 adultos en las habitaciones según las normas anteriores.

Nombre	Número de chicos	Número de chicas	Nombre del adulto
Roja			
Azul			
Verde			
Púrpura			
Naranja			
Amarilla			
Blanca			

Figura 3. Ítem de evaluación de la actuación

Para el problema *Campamento* se adjudicaban dos puntos para la respuesta correcta, que ha de satisfacer todas las condiciones planteadas en el enunciado. Se adjudicaba un punto para una respuesta parcialmente correcta en la cual se incumplen una o dos condiciones, si bien no cumplir la misma condición más de una vez se considera sólo como un incumplimiento. Se adjudicaba cero puntos para otras respuestas.

Este problema trata de cómo organizar el mejor itinerario para unas vacaciones. Las figuras 1 y 2 muestran un mapa de la zona y las distancias en kilómetros entre las ciudades.

Figura 1: Mapa de las carreteras que hay entre ciudades

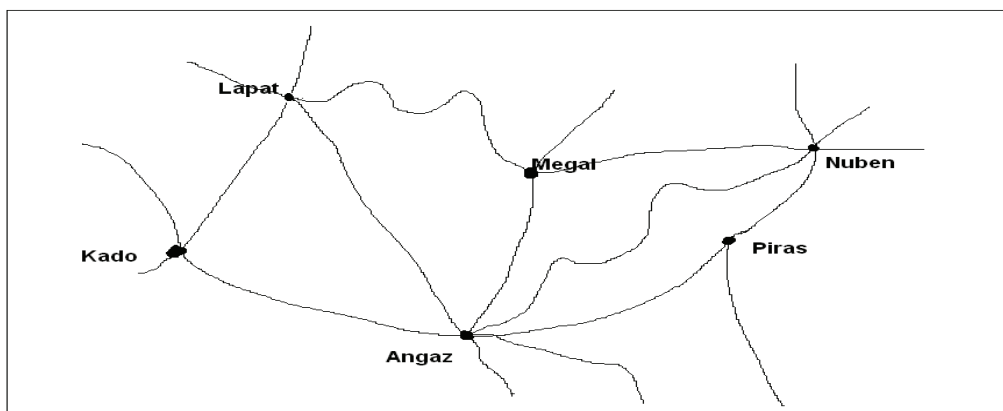


Figura 2: Distancias más cortas entre las ciudades en kilómetros

Angaz						
Kado	450					
Lapat	500	300				
Megal	300	-	550			
Nuben	550	-	-	450		
Piras	250	700	-	-	200	
	Angaz	Kado	Lapat	Megal	Nuben	Piras

¿Cuál es la distancia más corta por carretera entre Nuben y Kado? Exprese su respuesta en kilómetros. _____

Figura 4. Ítem de respuesta corta

Utilizando la información proporcionada en el enunciado de la pregunta anterior y suponiendo que estamos en Nuben, ¿cuál es el recorrido más corto por carretera para llegar a Lapat?

- Nuben - Angaz - Lapat
- Nuben - Piras - Angaz - Lapat
- Nuben - Megal - Lapat
- Nuben - Angaz - Kado - Lapat

Figura 5. Ítem de elección múltiple

Para el ítem de respuesta corta se considera *900 Km.* como la respuesta correcta y se le adjudicaba un punto, otras respuestas se consideraban incorrectas y se les adjudicaba cero puntos.

Para el ítem de elección múltiple se considera la opción *b* como la respuesta correcta y se le adjudicaba un punto, el resto de opciones se consideran respuestas incorrectas y se les adjudica cero puntos.

Con estas tres preguntas se elaboraron cuatro formas distintas de la prueba de competencias cuya diferencia radicaba en el orden de presentación de las preguntas; en la forma 1 el orden es EM, RC y EA, en la forma 2 RC, EM y EA, en la forma 3 EA, EM y RC y en la forma 4 EA, RC y EM. Con esta inversión en el orden de presentación de los ítems se buscaba controlar el efecto que éste pudiera tener en los examinados e intentar disminuir la posibilidad de copia entre los estudiantes. No se administraron formas que intercalaran el ítem con formato de evaluación de la actuación entre las preguntas de elección y respuesta corta debido a que estas últimas comparten el mismo enunciado.

El anexo A incluye un ejemplar de la forma 1. Como se puede observar, al comienzo de la misma el examinado debe proporcionar los datos personales que permitan identificarlo e información sobre otras variables de interés tales como sexo y fecha de nacimiento. Inmediatamente después aparecen las instrucciones en las que se le indica a los examinados que lean y resuelvan las preguntas que se presentan a continuación, identificadas con las letras A, B o C. El tiempo estimado para la ejecución de esta prueba fue de 20 minutos.

3. 1. 2. Prueba general

La prueba general fue elaborada con el objeto de reunir la mayor cantidad de información sobre un conjunto importante de variables del modelo, a saber las relacionadas con: (1) la percepción del sujeto sobre la tarea a realizar, (2) aspectos metacognitivos y (3) los componentes de ejecución o procesamiento cognitivo que se

activan para responder a las tres preguntas de la prueba de competencias.

Para la elaboración de la prueba general se partió del estudio de las teorías y modelos que sobre estas variables se han desarrollado. Se consideraron diferentes propuestas sobre motivación, autoeficacia, metacognición, estrategias cognitivas, conocimiento estratégico, componentes de ejecución y demás constructos relacionados con las variables de interés, revisándose un amplio conjunto de instrumentos elaborados a partir de estas teorías y modelos.

3. 1. 2. 1. Instrumentos revisados

En la Tabla 1 se ofrece una muestra de los instrumentos ofrecidos por la literatura especializada para medir características relacionadas con la tarea demandada al sujeto y con aspectos metacognitivos implicados en la misma.

Tabla 1. *Instrumentos revisados en relación a las variables incluidas en el modelo*

Instrumentos	Autor y año	Variables
<i>Multidimensional Scales of Perceived Self-Efficacy</i>	Bandura (1990, 2006)	Percepción sobre la tarea y sobre sí mismo
<i>Self-and-Task-Perception Questionnaire</i>	Eccles y Wigfield (1995)	
<i>Metacognitive Awareness Inventory</i>	Scraw y Dennison (1994)	Aspectos metacognitivos
<i>Motivated Strategies for Learning Questionnaire</i>	Pintrich, Smith, García y McKeachie (1991, 1993)	
<i>Goal Orientation and Learning Strategies Survey</i>	Dowson y McInerney (2004)	
<i>State Metacognitive Inventory</i>	O'Neil y Abedi (1996)	
<i>Inventory of Metacognitive Self-Regulation</i>	Howard, McGee, Shia y Hong (2000)	

El instrumento *Multidimensional Scales of Perceived Self-Efficacy* está basado en la teoría de la autoeficacia de Bandura (1995, 1997) y consta de 57 ítems que representan nueve dominios diferentes de autoeficacia: consecución de recursos

sociales, rendimiento académico, aprendizaje autoregulado, aprovechamiento del tiempo libre y de las actividades curriculares, eficacia autoreguladora para resistir a la presión de los compañeros, satisfacción de las expectativas de los otros, autoeficacia social, eficacia autoasertiva y consecución de ayuda de los padres y compañeros.

La escala *Self-and-Task-Percepcion Questionnaire*, basada en la teoría de expectativa-valor, tiene 19 ítems y considera dos dimensiones básicas: (1) expectativa de éxito con tres subescalas que son habilidad percibida, dificultad percibida de la tarea y esfuerzo requerido para realizar la tarea, y (2) valor percibido de la tarea con otras tres subescalas constituidas por interés en la tarea, utilidad de la tarea e importancia de la tarea.

Metacognitive Awareness Inventory es un cuestionario de 52 ítems diseñado para medir conocimiento de la cognición y regulación metacognitiva en el proceso de aprendizaje. Los 17 ítems indicadores de conocimiento de la cognición se distribuyen en los siguientes dominios: conocimiento declarativo, procedimental y condicional. Los 35 ítems indicadores de regulación metacognitiva se reparten en las siguientes categorías: planificación, estrategias de manejo de información, monitorización, estrategias para depurar errores y evaluación.

El cuestionario *Motivated Strategies for Learning Questionnaire* consta de 81 ítems distribuidos en dos secciones: motivación y estrategias de aprendizaje. La sección de motivación se compone de 31 ítems distribuidos en seis subescalas: orientación hacia metas intrínsecas, hacia metas extrínsecas, valor de la tarea, autoeficacia para el aprendizaje y la ejecución, creencias sobre control del aprendizaje y ansiedad. La sección de estrategias de aprendizaje tiene 50 ítems repartidos en nueve subescalas: práctica, elaboración, organización, pensamiento crítico, autoregulación metacognitiva, manejo del tiempo y del ambiente de estudio, regulación del esfuerzo, aprendizaje de

los compañeros y búsqueda de ayuda. Las 15 subescalas fueron diseñadas para ser usadas juntas o separadamente según las necesidades del investigador.

El instrumento *Goal Orientation and Learning Strategies Survey* es un instrumento diseñado para evaluar motivación para el logro de metas y uso de estrategias cognitivas y metacognitivas durante el proceso de aprendizaje. Está constituido por 84 ítems con los que se miden una selección de (1) metas académicas: dominio, ejecución y evitación del trabajo; (2) metas sociales: afiliación, aprobación, preocupación y responsabilidad social; (3) estrategias cognitivas: elaboración, organización y práctica, y (4) estrategias metacognitivas: monitorización, planificación y regulación.

El cuestionario *State Metacognitive Inventory* elaborado por O'Neil y Abedi (1996) consta de 20 ítems y considera cuatro subescalas para evaluar consciencia cognitiva, estrategias cognitivas, planificación y auto-chequeo. O'Neil y Brown (1998) realizaron una adaptación de este inventario para evaluar los procesos metacognitivos y afectivos que se activan en una situación de evaluación. Este nuevo cuestionario tiene 24 ítems y consta de cuatro subescalas de 6 ítems cada una: autochequeo, estrategias cognitivas, esfuerzo y preocupación.

El instrumento *Inventory of Metacognitive Self-Regulation* se desarrolló como una herramienta de medición útil para la evaluación y utilización del aprendizaje autoregulado en el aula de clase. Consta de 32 ítems distribuidos en cinco dimensiones, a saber conocimiento de la cognición, confianza en lo que se sabe, planificación, monitorización y evaluación.

La consistencia interna reportada para las escalas incluidas en la Tabla 1 está en un rango de 0.62 a 0.93. Todas estas escalas fueron objeto de análisis factoriales

confirmatorios para investigar la validez de las mismas y se encontró en todos los casos que la matriz estimada reproducía con bastante adecuación la matriz muestral.

Aún cuando estos instrumentos presentan propiedades psicométricas aceptables, ninguno de ellos resulta adecuado a las características particulares de este estudio. Por un lado, las escalas para medir variables asociadas a la percepción del sujeto sobre sí mismo o sobre la tarea están diseñadas apuntando hacia constructos muy específicos en razón de lo cual deben ser adaptadas a los requerimientos de cada investigador. Los instrumentos elaborados para evaluar regulación cognitiva y uso de estrategias cognitivas, en su mayoría, se han desarrollado con el propósito de evaluar estas variables dentro del contexto de las estrategias para el aprendizaje y no se orientan a evaluar el comportamiento del sujeto en la ejecución de tareas concretas. Por otra parte, aun en el caso de que pudieran adaptarse algunos de ellos, está la dificultad añadida que plantea la longitud de la mayoría de estos instrumentos, que hace inviable su administración en el tiempo de aplicación disponible en el marco de trabajo de un centro escolar, donde es difícil conseguir la cesión de sesiones más allá de una hora u hora y media.

Por tanto, se estimó necesario proceder a la redacción de un conjunto de ítems que permitieran estructurar un único cuestionario ajustado a los requerimientos de la investigación, partiendo de la revisión teórica y la investigación previa.

3. 1. 2. 2. Redacción de los ítems

Para la prueba general se especificaron seis variables a medir: las relacionadas con la percepción del sujeto sobre la tarea (*valor percibido de la tarea y dificultad percibida de la tarea*), las concernientes al ámbito metacognitivo (*conocimiento autopercebido, regulación metacognitiva y conocimiento estratégico*) y dentro del ámbito cognitivo los *componentes de procesamiento cognitivo*.

Se comenzó redactando los ítems que servirían como indicadores de la percepción del sujeto sobre la tarea y sobre sí mismo puesto que para ambas variables la elaboración de los ítems estuvo basada en las categorías señaladas en la teoría de la autoeficacia de Bandura (1995, 1997, 2006) y en la teoría de expectativa-valor de Eccles y cols. (1983, 1994, 1995). Los ítems eran redactados de modo que el examinado respondiera juzgando tanto el contenido de la tarea como el formato en el que ésta se presentaba.

El valor percibido de la tarea se expresa en función de:

- su interés o valor intrínseco, definido como el disfrute que experimenta un sujeto cuando realiza una tarea o como el interés subjetivo en su contenido;
- la importancia de la tarea, definida como la trascendencia que se le da al hecho de realizar bien la tarea; y
- su valor utilitario o extrínseco, que se define como la utilidad de la tarea para el individuo en términos de sus metas futuras.

Los ítems utilizados como indicadores de esta variable son los que se muestran en la Tabla 2.

Tabla 2. *Ítems de la prueba general para medir el valor percibido de la tarea*

Dimensión	Ítem	Nº del ítem
Valor intrínseco	La tarea planteada en esta pregunta es entretenida	3
	Me gusta la tarea que se plantea en la pregunta	6
	Las preguntas planteadas con este formato resultan tediosas	10
	Si pudiera elegir un tipo particular de examen, me gustaría que las preguntas tuvieran este formato	12
Importancia	Esta pregunta permite demostrar lo que uno sabe	4
	Es importante saber responder a preguntas formuladas de este modo	11
Valor extrínseco	La tarea que plantea la pregunta es de poca utilidad para mi formación futura	5
	Estar familiarizado con preguntas planteadas con este tipo de formato puede ser de utilidad en el futuro	13

La *dificultad percibida de la tarea* está relacionada con los juicios del examinado sobre el grado de dificultad y el nivel de esfuerzo que pueda demandar la tarea que debe ejecutar. Los ítems que se elaboraron para la medición de esta variable se presentan en la Tabla 3.

Tabla 3. *Ítems de la prueba general para medir la dificultad percibida de la tarea*

Dimensión	Ítem	Nº del ítem
Dificultad de la tarea	La pregunta me pareció difícil	7
	El tipo de formato utilizado hace que la pregunta parezca más difícil	16
Esfuerzo requerido	Para contestar correctamente esta pregunta es necesario saber mucho	8
	Con preguntas planteadas con este formato es fácil aprobar un examen	14
	Las preguntas planteadas con este formato exigen mucho esfuerzo para ser contestadas correctamente	15

Para estimar la variable *conocimiento autopercebido* o percepción sobre sí mismo se elaboraron ítems que permitieran conocer los juicios que el examinado hace sobre su competencia en la ejecución de la tarea y que representan lo que él cree que es capaz de hacer, más no necesariamente sus competencias reales (véase Tabla 4).

Tabla 4. *Ítems de la prueba general para medir conocimiento autopercebido*

Ítem	Nº del ítem
Creo que contesté correctamente la pregunta	1
Se lo suficiente como para responder correctamente a la pregunta	2
Me resulta sencillo responder a preguntas con este tipo de formato	9

Para los ítems correspondientes a la variable *regulación metacognitiva* se consideraron como actividades básicas la planificación y el control, que incluye tanto los procesos de monitorización y de evaluación (Brown, 1987; Dowson y McInerney, 1998, 2004; García y McKeachie, 2005; Justice y Dornan, 2001; Meijer, Veenman y van Hout-Wolters, 2006; Pintrich y deGroot, 1990; Pintrich et al., 1993), de manera que se redactaron los ítems apuntando a determinar el empleo de estas destrezas metacognitivas durante la ejecución de las tareas planteadas en la prueba de

competencias. Se consideró la planificación como la actividad que el sujeto ha de hacer antes de comenzar la tarea y que debe traducirse en una esmerada preparación para completarla satisfactoriamente, que involucra entender lo que se pide, identificar, extraer y priorizar la información relevante, establecer las metas y decidir qué pasos y recursos usar para resolver el problema (Davidson, Deuser y Sternberg, 1994; Dowson y McInerney, 2004). El control se refiere a la implementación de medidas de autochequeo, dirigidas a la observación del funcionamiento de las estrategias y de los progresos en la ejecución de la tarea con miras al logro de la meta y a la evaluación de la calidad de la solución, todo lo cual involucra decisiones sobre el uso de nuevas estrategias o la adaptación de las utilizadas a fin de contrarrestar dificultades o deficiencias identificadas durante la monitorización (Davidson, Deuser y Sternberg, 1994; Dowson y McInerney, 2004; Kluwe, 1987). Los ítems que se muestran en la Tabla 5 representan los indicadores propuestos en este estudio para determinar el uso de estos procesos de regulación metacognitiva.

Tabla 5. *Ítems de la prueba general para medir regulación metacognitiva*

Actividad	Ítem	Nº del ítem
Planificación	<i>Antes de responder:</i>	
	¿ha leído detenidamente el enunciado de la pregunta?	17
	¿se ha asegurado de haber entendido qué es lo que se pedía?	18
	¿ha reformulado el enunciado con sus propias palabras para poder comprender mejor lo que se pedía?	19
	¿ha buscado y entresacado la información principal proporcionada por el enunciado?	20
	¿ha hecho un pequeño plano de situación con la información proporcionada y con lo que se pedía?	21
	¿se ha planteado alguna estrategia concreta para realizar la tarea requerida?	22
	¿sabía perfectamente lo que tenía que hacer para llegar a la respuesta correcta?	23
Control	<i>Conforme iba respondiendo</i>	
	¿ha ido chequeando su trabajo?	24
	¿ha encontrado alguna dificultad al realizar la tarea demandada?	25
	<i>¿Qué ha hecho para intentar resolver esa dificultad?</i>	
	Seguir adelante como si nada	26
	Parar y leer de nuevo el enunciado de la pregunta	27
	Buscar una nueva manera de resolver la pregunta	28
	<i>Una vez respondida la pregunta</i>	
	¿revisó su trabajo para asegurarse de no tener errores?	37

Para la elaboración del conjunto de ítems que constituyen los indicadores de *conocimiento estratégico* (véase Tabla 6) se consideraron las estrategias o métodos heurísticos que podrían ser utilizados por los estudiantes en la solución de los problemas propuestos en la prueba de competencias. Entre los heurísticos que se consideraron están el trabajo en sentido inverso, el planteamiento de problemas análogos, la descomposición de la tarea en sub-metas y el uso de estrategias mixtas (Bruning, Schraw, Norbu y Ronning, 2005; Fan & Zhu, 2007; Nickerson, 1994).

Tabla 6. *Ítems de la prueba general para medir conocimiento estratégico*

Ítem	Nº del ítem
¿Consideró varias formas de trabajo para responder la pregunta?	29
¿Ha tratado de dividir el problema en pequeños pasos?	30
¿Planteó una situación similar para comprender la pregunta formulada y poder responderla?	31
Antes de hacer cálculos, ¿descartó las alternativas que no le parecían respuestas posibles?	32
¿Utilizó procedimientos que aunque no conducían directamente a la solución le ayudaron a obtener la respuesta?	33
¿Respondió la pregunta partiendo de una respuesta probable?	34
¿Trabajó con una estrategia que ha utilizado en otras ocasiones en tareas similares?	35
¿Su forma de trabajo ha sido guiada por el análisis lógico del problema planteado en la pregunta?	36

Para el examen de los *componentes de procesamiento cognitivo* puestos en marcha en la solución de las tareas de la prueba de competencias se optó por el planteamiento de Sternberg. Como se ha señalado previamente, los componentes de procesamiento de la teoría triárquica se organizan en tres fases que son codificación, combinación y respuesta. En el caso que nos ocupa estos componentes básicamente serían comprender el significado del problema planteado, buscar y ubicar la información necesaria, organizar esa información, hacer uso de la información para el logro de una respuesta, revisar la corrección de la solución y emitir una respuesta. Los ítems elaborados permiten conocer el grado de dificultad que supone el empleo de cada componente de procesamiento cognitivo para el examinado (véase Tabla 7).

Tabla 7. *Ítems de la prueba general para medir componentes de procesamiento cognitivo*

Fase	Ítem	Nº del ítem
Codificación	Comprender con claridad lo que pide el problema	38
	Identificar las variables implicadas en la solución	39
	Identificar las condiciones que debe cumplir la solución	40
Combinación	Organizar la información proporcionada	41
	Hacer cálculos	42
	Comprobar la adecuación de la solución	43
Respuesta	Seleccionar o escribir la respuesta correcta	44

Suplementariamente a este conjunto de ítems se planteó utilizar la técnica de los reportes o protocolos verbales mediante la cual el estudiante debía hacer una descripción oral con el mayor grado posible de detalle de lo que había hecho para llegar a la solución propuesta. Se consideró adecuado utilizar esta técnica con el fin de tener información adicional en la identificación de estos componentes clave así como de las estrategias utilizadas para resolver la tarea y también para examinar la coherencia entre el discurso del sujeto y lo indicado en la prueba.

Para llegar a los ítems definitivos, luego de su redacción preliminar, cada grupo inicialmente elaborado era exhaustivamente revisado y a partir de este análisis, en caso de ser necesario, se procedía a su re-elaboración, eliminando algunos ítems, introduciendo o re-escribiendo otros, para así llegar a la prueba general conformada por las 44 preguntas recogidas en las tablas anteriores. Todas las preguntas debían ser respondidas por el estudiante en relación a cada una de las tareas que había realizado en la prueba de competencias, a fin de tener información sobre todas las variables para cada uno de los formatos estudiados.

3. 1. 2. 3. Confección de la prueba

La prueba general (véase Anexo B) contiene los siguientes elementos:

- Un apartado inicial para datos personales dado que, aún cuando se aplicaría en la misma ocasión que la prueba de competencias, se implementaría separadamente.

- Un recuadro de instrucciones generales adaptadas a la forma correspondiente de la prueba de competencias en función del formato de ítem asociado a cada una de las preguntas A, B o C.
- Las indicaciones particulares para cada apartado, en alguno de los cuales se presenta un ejemplo de cómo debería procederse para emitir la respuesta.

Los indicadores correspondientes a cada variable se organizaron en cinco partes. Las dos primeras corresponden a los indicadores de percepción sobre sí mismo, valor y dificultad percibida de la tarea, distribuidos en cada apartado en función de si el ítem se orienta a medir lo relativo al contenido de la tarea (1^{er} bloque) o lo relativo al formato en que ésta se presenta (2^o bloque); cada parte consta de 8 ítems con formato Likert con cuatro opciones de respuesta. La tercera parte contiene los 12 ítems que se constituyen como indicadores de los procesos de regulación metacognitiva, presentados con formato dicotómico en función de si el sujeto llevó a cabo la acción señalada en el ítem. La cuarta parte incluye los indicadores de conocimiento estratégico, conformada por 9 ítems con formato de respuesta dicotómica en los que se recoge la respuesta del examinado sobre si usó o no las estrategias o heurísticos que se le señalan. La quinta y última parte agrupa los indicadores que se ocupan de los componentes de procesamiento y tiene 7 ítems con un formato de respuesta con cuatro opciones que representan el nivel de dificultad encontrado por el sujeto en las operaciones implicadas en cada uno de los componentes de procesamiento y una quinta opción señalada como “no procede” para el caso de que el examinado considere que no realizó ese procedimiento o actividad al responder al ítem.

La distribución de las preguntas y de cada una de las partes de esta prueba se hizo de modo que la prueba resultara una secuencia organizada y lógica, tratando de evitar al máximo que la respuesta a un ítem pudiera ejercer alguna influencia en la

próxima respuesta; cómoda de responder, agrupando los ítems según la naturaleza de la variable a medir, y tratando de evitar la sensación de que el cuestionario podía ser muy largo, para lo cual en un mismo ítem se respondería acerca de las tres preguntas de la prueba de competencias.

Se cuidó particularmente la impresión, atendiendo a que todos los elementos de los diferentes cuestionarios fueran perfectamente nítidos y legibles. Se utilizó un papel de tamaño superior al DIN A4, conocido como tamaño oficio, cuyas dimensiones son 22 cm. de ancho por 32 cm. de largo. Se estimó que 35 minutos sería el tiempo de aplicación adecuado para que los examinados completaran la prueba general.

Una docente venezolana de amplia experiencia revisó el contenido de las pruebas con el objeto de corroborar que la redacción era clara y de fácil comprensión para los estudiantes que serían examinados y, en caso necesario, hacer las adaptaciones pertinentes en la redacción de las mismas.

3. 1. 3. Hoja de registro de los profesores

Para obtener información acerca del nivel de conocimiento declarativo y procedimental de los estudiantes hubo que solicitar una valoración cuantitativa a sus profesores de Matemáticas, ya que la información procedente de la prueba de competencias, formada solo por tres ítems, no podía proporcionar una estimación adecuada de los mismos.

A cada profesor se le suministraba una hoja de registro (véase Anexo C) constituida por un listado con los nombres de los estudiantes examinados correspondientes a su grupo y tres columnas, una para registrar el nivel de conocimiento declarativo, otra para el conocimiento procedimental y otra para la última calificación en matemáticas de cada uno de esos estudiantes como una medida de su dominio de las

matemáticas, considerado por algunos investigadores como una variable importante en la solución de problemas (Fitzpatrick, 1994; Putnam, 1987).

3. 2. SUJETOS

Se aplicaron las pruebas a una muestra de conveniencia formada por 556 participantes, estudiantes de 4º curso de bachillerato (equivalente a 4º de ESO) de siete centros educativos ubicados en el Estado Trujillo en Venezuela. La selección de los estudiantes de 4º curso obedeció al hecho de que la mayoría de los estudiantes de este curso tienen 15 años, que es la edad para la que están diseñadas las preguntas del estudio PISA utilizadas en la prueba de competencias. Las Figuras 6, 7 y 8 indican la distribución muestral por edad, sexo y titularidad de los centros escolares.

Se observa en la Figura 6 una distribución equitativa en función del sexo de los examinados que conforman la muestra (50.2 % masculino y 49.8% femenino). En la Figura 7 se advierte que la mayoría de los examinados tiene una edad de 15 años (73.7 %) y el resto con edades muy próximas. En la Figura 8 se observan porcentajes relativamente similares en cuanto a titularidad de los centros educativos: centros privados (42.6 %) y centros públicos (57.4 %).

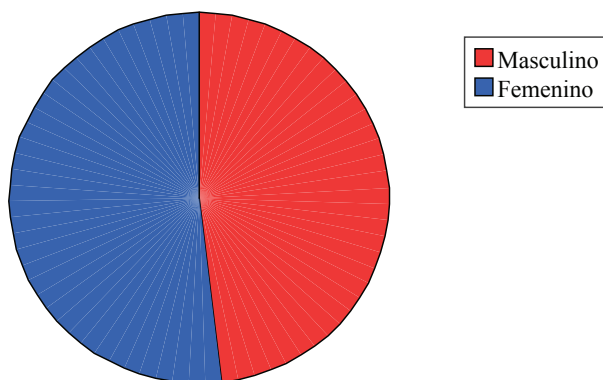


Figura 6. Distribución por sexo

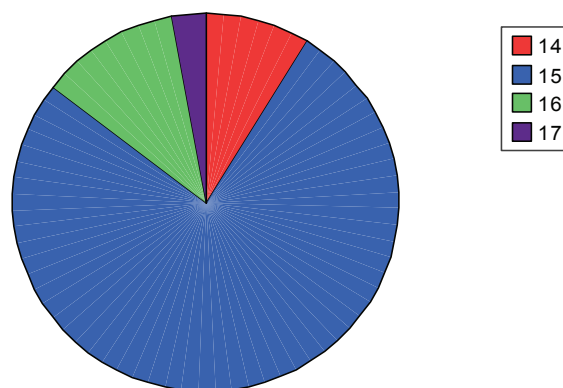


Figura 7. Distribución por edad

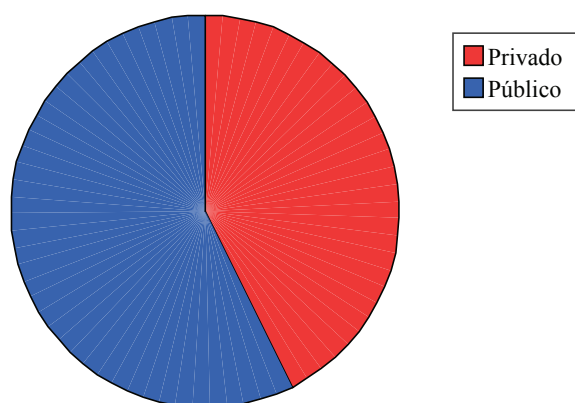


Figura 8. Distribución por titularidad del centro

3. 3. PROCEDIMIENTO

La recogida de datos se desarrolló en Venezuela durante tres semanas, desde el 15 de octubre hasta el 2 de noviembre de 2007. Las fases planificadas para el desarrollo del trabajo de campo fueron las siguientes:

1. Contacto con los centros escolares
2. Preparación del material para la aplicación
3. Formación de los aplicadores

4. Aplicación piloto
5. Aplicación en el aula
6. Recogida de los protocolos verbales
7. Recogida de la información de los profesores

3.3.1. Contacto con los centros escolares

Se procedió a contactar con un conjunto de 12 institutos educativos del Estado Trujillo en los que se mantuvo una entrevista personal con sus directores o coordinadores académicos.

Durante esta entrevista se informaba de manera general de las características de la investigación en curso y se explicaba en detalle en qué consistía toda la colaboración que se requería, tanto de los estudiantes como de los profesores, a fin de que pudieran valorar con exactitud sus posibilidades de participar en el estudio. Al finalizar la entrevista se les hacía entrega de una carta invitándoles formalmente a participar en la investigación y un resumen del proyecto (véase Anexo D).

Todos los directores o coordinadores académicos dieron directamente su respuesta durante la entrevista, negativa en cinco casos y positiva en siete institutos (cuatro de titularidad privada y tres de titularidad pública).

En esa misma reunión se designó la persona de contacto entre el instituto y los responsables de la investigación y se celebró ese mismo día la primera reunión de trabajo, con el fin de (1) obtener información acerca del número de grupos en 4º curso y el número de alumnos por grupo, (2) fijar la fecha y hora para la aplicación de las pruebas y (3) determinar la sala para la recogida de datos mediante protocolos verbales.

3.3.2. Preparación de los materiales

Una vez determinado el número de estudiantes que iba a participar en el estudio, se procedió a la impresión de las dos pruebas que debían ser completadas por éstos durante la aplicación: la prueba de competencias, con los tres ítems de diferente formato y la prueba general.

Completado el proceso de impresión se procedió a ensamblar el material, colocando cada una de las cuatro formas de la prueba de competencias con su correspondiente forma de la prueba general. Para las primeras aplicaciones se prepararon paquetes de 25 juegos ordenados secuencialmente de la forma 1 a la forma 4 hasta completar los 25 ejemplares previstos inicialmente para cada grupo de estudiantes a evaluar en un aula, con el objeto de facilitar la distribución en forma de espiral de las pruebas en el aula donde se realizaría la administración y así garantizar que tendríamos una asignación aleatoria entre los examinados y una repartición proporcional de las 4 formas de la prueba de competencias.

3.3.3. Formación de los aplicadores

Con el fin de estandarizar el procedimiento de recogida de datos en los centros, se organizó una sesión de formación con las personas que colaborarían como aplicadores. El propósito de esta reunión fue explicarles el objetivo de la investigación, examinar en detalle el contenido de los cuestionarios y, sobre todo, mostrar cómo había que hacer la aplicación.

En esta sesión se les indicó que la aplicación debía realizarse siguiendo escrupulosamente las pautas indicadas en las instrucciones elaboradas a tal efecto que aparecen en el *Guión para el aplicador del test* (véase Anexo E), el cual recoge de manera pormenorizada todo el trabajo a desarrollar durante la aplicación y especifica el

texto concreto de las intervenciones del examinador durante la administración de las pruebas. Junto a este documento se les entregó también un resumen del proyecto idéntico al entregado a los directores de los institutos educativos. Durante esta sesión se acordó con cada aplicador el calendario de aplicaciones que le correspondía cumplir.

3.3.4. Aplicación piloto

Se llevó a cabo en la Unidad Educativa (U. E.) Monseñor Vicente Valera Márquez, de titularidad privada, con dos grupos a los que se les administraron las pruebas simultáneamente y en un mismo salón a fin de permitir la supervisión de la doctoranda. El objetivo era examinar aspectos relativos al tiempo de ejecución de los cuestionarios, al funcionamiento del *Guión para el aplicador del test* y comportamiento general del estudiantado ante la administración de las pruebas.

A partir de la experiencia recabada en esta aplicación piloto se observó que el tiempo de aplicación de la prueba de competencias establecido inicialmente en 20 minutos resultaba insuficiente y que ampliarlo a 30 minutos no introducía inconvenientes en la administración de las pruebas en razón de que aún así la aplicación no excedía el tiempo disponible cedido por los institutos (un módulo de 90 minutos).

Se observó también que el protocolo de actuación recogido en el guión funcionaba de modo correcto. Sin embargo, al establecerse en el mismo que el aplicador debía leer las instrucciones de la prueba general se introducía un elemento de confusión entre los estudiantes, debido a que el orden de presentación de las tres preguntas de la prueba de competencias es distinto para cada una de las cuatro formas y en las instrucciones se hace mención a cada una de estas tres preguntas asociándolas con la letra que le corresponde en su respectiva forma. Aún así se consideró imprescindible la lectura de las instrucciones generales y se organizó la distribución de

las pruebas de manera diferente: distintas formas entre grupos e igual forma dentro de los grupos.

También pudo advertirse que la presencia de un miembro de la planta profesoral del centro escolar contribuía enormemente en el mantenimiento de la disciplina estudiantil.

En base a la información que se obtuvo de la aplicación piloto se introdujeron los cambios necesarios en el guión para el aplicador de los cuales fueron instruidos oportunamente a fin de que los conocieran previamente a sus aplicaciones y los implementaran debidamente. En cuanto a los cuestionarios, se procedió a empaquetar el material haciendo sencillamente bloques de 25 ejemplares correspondientes todos ellos a la misma forma de las pruebas.

3.3.5. Aplicación en el aula

La persona de contacto en el centro escolar se ocupaba de hacer una presentación de los responsables de la investigación y solicitar la máxima colaboración a los estudiantes, que consistiría en participar voluntariamente en las pruebas que a continuación se aplicarían. La doctoranda invitaba entonces a los estudiantes que quisieran a colaborar con los protocolos verbales explicándoles en qué consistiría su tarea y acto seguido el examinador procedía de acuerdo al protocolo establecido en el *Guión para el aplicador del test*. Para la prueba de competencias tenían un tiempo de ejecución máximo de 30 minutos, pasados los cuales el examinador pasaba por los pupitres a recogerlas y a continuación debían contestar la prueba general en un tiempo de 35 minutos. Una vez recogida la prueba general se agradecía a los estudiantes su participación y se daba por concluida la aplicación. En cada aplicación se contó con la presencia de un miembro de la planta profesoral del centro escolar.

Aún cuando ciertas complicaciones ajenas a la investigación impidieron que se realizaran la totalidad de las aplicaciones planificadas, no se presentaron problemas importantes durante la administración de los cuestionarios. El número de grupos planificados, el número de grupos y de estudiantes examinados en cada centro escolar se muestra en la Tabla 8.

Tal y como se observa en la Tabla 8, el número de grupos planificados difiere del número de grupos examinados en cuatro de los siete institutos que participaron en el estudio. Esto obedeció, como ya se mencionara, a causas de diversa índole (suspensión intempestiva de actividades escolares, cambios de última hora en la planificación por parte de los encargados del plantel, entre otras) que no pudieron ser controladas por el investigador. Estas situaciones no sólo impidieron realizar todas las aplicaciones pautadas, incidiendo en el hecho de no poder obtener un número equilibrado de estudiantes por cuadernillo, sino que obligaron a cambiar el cronograma de las aplicaciones (simultáneamente o inmediatamente después una de la otra) para algunos grupos, con el consiguiente desmedro de la confidencialidad de las preguntas de la prueba, que obligó a ser muy cauto en los análisis preliminares de los resultados obtenidos.

Tabla 8. *Relación de centros escolares que participaron en el estudio*

CENTRO ESCOLAR	Nº de grupos planificados	Nº de grupos examinados	Nº de sujetos examinados
Liceo Bolivariano Monay	8	6	162
U. E. Ntra. Sra. De La Paz	3	3	88
U. E. Mons. Vicente Valera Márquez	2	2	55
U. E. Colegio Ntra. Sra. De Coromoto	2	2	64
U.E. Colegio Los Cedros	3	1	30
Liceo Bolivariano Cristóbal Mendoza	16	5	121
Liceo Bolivariano Pedro J. Carrillo M.	3	2	36
TOTAL	37	21	556

3. 3. 6. Recogida de los protocolos verbales

Como se ha indicado en el apartado anterior, la invitación a los estudiantes a colaborar con los protocolos verbales se realizaba una vez que la persona de contacto presentaba a los responsables de la investigación. Se les explicaba que debían responder a la prueba de competencias y a continuación relatar oralmente el procedimiento seguido para responder a la pregunta, lo cual sería registrado mediante una grabación. Se obtuvo una respuesta prácticamente nula y en consecuencia la recolección de los mismos no pudo llevarse a cabo. Los estudiantes se negaron rotundamente a colaborar, aduciendo como razón principal a esta negativa un problema de miedo escénico ante el grabador y en segundo lugar, que tenían dificultad para expresar adecuadamente los procedimientos mentales seguidos para realizar la tarea. Sólo se logró obtener dos reportes verbales que resultaron de poco valor para ser considerados en la investigación.

Es necesario señalar que, con el objeto de encargarse de la recogida del protocolo verbal, la doctoranda acompañaba al aplicador designado en cada grupo, pero al no poder concretarse esta tarea, la doctoranda se ocupaba entonces de la aplicación, con excepción de aquellos casos en los que se producía la aplicación de modo simultáneo en dos grupos o más. De este modo, de las personas entrenadas para actuar como aplicadores sólo cuatro de las seis colaboraron en la aplicación y en la mayoría de los casos las pruebas fueron aplicadas por la propia doctoranda.

3. 3. 7. Recogida de la información de los profesores

La hoja de registro para los profesores de matemáticas de los distintos grupos involucrados en el estudio se les hizo llegar posteriormente a la aplicación de las pruebas a los estudiantes a través de la persona de contacto del centro, quien se encargó de su distribución y recolección en los centros escolares, así como de la devolución del

material, una vez completado por los profesores, a la persona encargada por la responsable de la investigación. Conjuntamente con el listado de su(s) grupo(s) se le dio a cada profesor de matemáticas una carta explicativa de su tarea junto con el material informativo necesario que le permitiera hacer la mejor estimación posible sobre el nivel de conocimiento declarativo y procedimental de los alumnos participantes en el estudio (véase Anexo F). Este material estaba constituido por un cuadro explicativo del significado de cada tipo de conocimiento, un ejemplar de la prueba de competencias con la que habían sido examinados los sujetos y el resumen del proyecto de investigación. La recepción de este material se produjo durante el lapso comprendido entre los meses de Enero y Marzo de 2008 con una tasa de respuesta del 100%.

3. 4. PLAN DE ANÁLISIS

Previo al inicio de todos los análisis se procedió a una revisión del fichero de datos, examinando una muestra aleatoria del 10% de los casos de la que se verificó un porcentaje de error del 0 % en la codificación y grabación de los datos, realizadas por la doctoranda.

Luego de confirmado este extremo se procedió al análisis de los datos, el cual se desarrolló en varias etapas. Se comenzó con un estudio para el control de calidad de los datos reunidos y se analizó la posible estructura jerárquica de los datos. A continuación, se construyeron las variables predictoras del modelo teórico formulado para finalmente estimar dicho modelo con los datos obtenidos.

3. 4. 1. Control de calidad

Se procedió a un exhaustivo análisis exploratorio y descriptivo de los datos

recogidos para examinar posibles inconsistencias y detectar la existencia de errores, con el fin de poderlos corregir y depurar la base de datos antes de comenzar los análisis conducentes a dar cuenta de las preguntas de investigación planteadas en el estudio. Además, este análisis posibilitó tener una primera aproximación a las características y comportamiento de la muestra reunida en relación a los aspectos de mayor interés.

En particular, se llevó a cabo un estudio univariado y bivariado de los datos reunidos, prestando especial atención a la presencia de patrones anómalos de respuesta, de no respuesta o valores perdidos y a la calidad métrica de las dos pruebas utilizadas en el estudio.

3. 4. 2. Análisis acerca de la estructura de los datos

El siguiente paso en el análisis fue examinar el tipo o estructura de los datos, con el fin de evaluar su independencia y, de este modo, tomar decisiones acerca de cuál sería la mejor herramienta para estimar el modelo propuesto teóricamente con los datos obtenidos.

En principio, dado que los estudiantes están anidados dentro de grupos y éstos dentro de centros, se consideró que el centro en el que estudia el alumno podría tener un efecto en su actuación en la prueba y, de ser así, sería necesario recurrir a una estrategia multinivel para modelar los datos. Para ello, se realizó un análisis de varianza de un factor de efectos aleatorios.

3. 4. 3. Construcción y análisis preliminar de las variables predictoras

En principio se construyeron las variables con arreglo al modelo propuesto y de seguida se sometieron a algunos análisis preliminares con la finalidad de ahondar en el estudio de los datos reunidos e ir en la vía de descubrir particularidades de cada

formato. Específicamente, se estimaron los valores descriptivos para cada una de ellas y se realizaron comparaciones de medias entre las distintas variables según formato.

3. 4. 3. 1. Construcción de las variables predictoras

La variable respuesta está representada por las respuestas de los sujetos a los ítems de la prueba de competencia (tres preguntas cada una con un formato diferente) y las variables predictoras o independientes serían cada una de las correspondientes a la medida de los factores que se plantean en el modelo como posibles determinantes de la respuesta al ítem, a saber:

- Dentro del ámbito cognitivo, el conocimiento declarativo (CD), el conocimiento procedimental (CP), la calificación en matemáticas (CM) y los componentes de procesamiento cognitivo (codificación, combinación y respuesta).
- Dentro del ámbito metacognitivo, el conocimiento autopercebido, el conocimiento estratégico y la regulación metacognitiva.
- Dentro del ámbito relativo a la tarea, el valor y la dificultad percibidas en la misma.

3. 4. 3. 1. 1. Variables del ámbito cognitivo

Las medidas de conocimiento declarativo y procedimental que se utilizarán serán las valoraciones que hicieran de estas variables los profesores de matemáticas de los estudiantes examinados, que también proporcionaron la calificación en matemáticas de estos estudiantes en la hoja de registro elaborada a tal efecto.

Para los componentes de procesamiento se construyó un conjunto de tres variables correspondientes a las fases señaladas en la sub-teoría componencial de Sternberg (1985, 1998a, 1998b):

- Codificación: engloba los componentes implicados en la comprensión del problema, constituyéndose la variable como la suma de las puntuaciones en los

ítems 38, 39 y 40 de la prueba general.

- Combinación: engloba a los componentes implicados en la solución del problema y está representada por la suma de las puntuaciones en los ítems 41, 42 y 43.
- Respuesta: se refiere al componente implicado en la comunicación de la respuesta y corresponde a la puntuación en ítem 44.

3. 4. 3. 1. 2. Variables del ámbito metacognitivo

Se consideró la variable conocimiento auto-percibido o percepción sobre sí mismo como la suma de las puntuaciones a los ítems 1, 2 y 9.

Para el conocimiento estratégico la variable quedó constituida por el número de estrategias utilizadas para realizar la tarea, presentadas en los ítems 30 al 36.

La regulación cognitiva se midió a través de dos procesos clave: planificación y control.

La variable planificación se organiza en función de lo que se hace en ese proceso:

- Asegurarse de comprender: 1 punto si el sujeto respondió afirmativamente al menos uno de los ítems del 17 al 19.
- Buscar y organizar información: 1 punto si respondió afirmativamente el ítem 20 o el ítem 21.
- Trazar un plan de trabajo: 1 punto si respondió afirmativamente el ítem 22 o el ítem 23.

Luego, la calificación para planificación es la suma de estas tres partes del proceso.

La variable control se organiza en función de lo que se hace en los procesos de monitorización y evaluación.

- Monitorización de su trabajo: un punto si respondió afirmativamente al ítem 24.
- Aplicación de acciones correctivas: cero punto si no respondió afirmativamente ningún ítem del 26 al 28 o respondió negativamente al ítem 25 y la hizo mal; un punto si respondió afirmativamente al menos un ítem del 26 al 28 y dos puntos si respondió afirmativamente al ítem 26 o al 27 y al 28 o respondió negativamente al ítem 25 y la hizo bien.
- Evaluación final: un punto si respondió afirmativamente al ítem 37

Luego, la calificación en la variable control se obtiene a partir de la suma de las puntuaciones en estas tres actividades del proceso.

3. 4. 3. 1. 3. Variables del ámbito relativo a la tarea

Para estas variables se optó por sumar las respuestas a los ítems correspondientes a cada dimensión:

- Valor percibido de la tarea en relación al contenido: ítems 3, 4, 5 y 6.
- Valor percibido de la tarea en relación al formato: ítems 10, 11, 12 y 13.
- Dificultad percibida de la tarea en relación al contenido: ítems 7 y 8.
- Dificultad percibida de la tarea en relación al formato: ítems 14, 15 y 16.

Tabla 9. Cuadro resumen con los ítems indicadores de cada variable predictora

Variable predictora	Ítems
Codificación (CODI)	38, 39 y 40
Combinación (COMB)	41, 42 y 43
Respuesta (RESP)	44
Conocimiento auto-percibido (CAP)	1, 2 y 9
Conocimiento estratégico (CES)	30 al 36
Planificación (PLAN)	17 al 23
Control (CONT)	24 al 28 y 37
Valor percibido de la tarea en relación al contenido (VTC)	3, 4, 5 y 6
Valor percibido de la tarea en relación al formato (VTF)	10, 11, 12 y 13
Dificultad percibida de la tarea en relación al contenido (DTC)	7 y 8
Dificultad percibida de la tarea en relación al formato (DTF)	14, 15 y 16

En la Tabla 9 se muestra de modo resumido la información relativa a los ítems que están involucrados en la construcción de cada variable predictora del modelo.

3. 4. 3. 2. Análisis preliminar de las variables predictoras

Se llevó a cabo un detallado estudio descriptivo de las nuevas variables creadas, con el fin de ver su distribución en la muestra del estudio y las posibles diferencias en los distintos formatos utilizados en la prueba de competencias, así como su relación entre ellas y con un buen número de los aspectos evaluados en el estudio.

3. 4. 4. Estimación del modelo

Una vez realizadas todas las tareas previas de revisión y preparación de los datos para la prueba empírica del modelo se ejecutó el modelado de las variables a objeto de poder dar respuesta a las preguntas que se plantea esta investigación.

Para someter a comprobación empírica el modelo teórico formulado, se utilizó un modelo lineal generalizado con el logit como función de enlace. Dado que para los ítems de EM y RC solo existen dos categorías de respuesta (correcta o incorrecta), se ha trabajado con la regresión logística binaria. El ítem de EA cuenta con tres categorías de respuesta posibles (correcta, parcialmente correcta, incorrecta), por lo que se ha de operar con la regresión logística ordinal.

Para estimar el modelo para cada uno de los formatos se trabajó siguiendo el enfoque descrito por Ato y cols. (2005, p. 57) que, con algunas modificaciones se reproduce en la Figura 9.

3. 4. 4. 1. Especificación

En esta fase se formula el componente sistemático del modelo, definido por el conjunto de variables predictoras del modelo teórico señaladas por la literatura y construidas según se indicó en el apartado anterior. Además del criterio teórico, en esta

fase se introdujo un criterio estadístico para decidir acerca de la inclusión de estos posibles predictores, realizando un análisis de regresión logística para examinar por separado la relación entre cada predictor y la variable de respuesta, considerando como criterio de inclusión una probabilidad asociada al coeficiente de regresión del predictor menor o igual a 0.05.

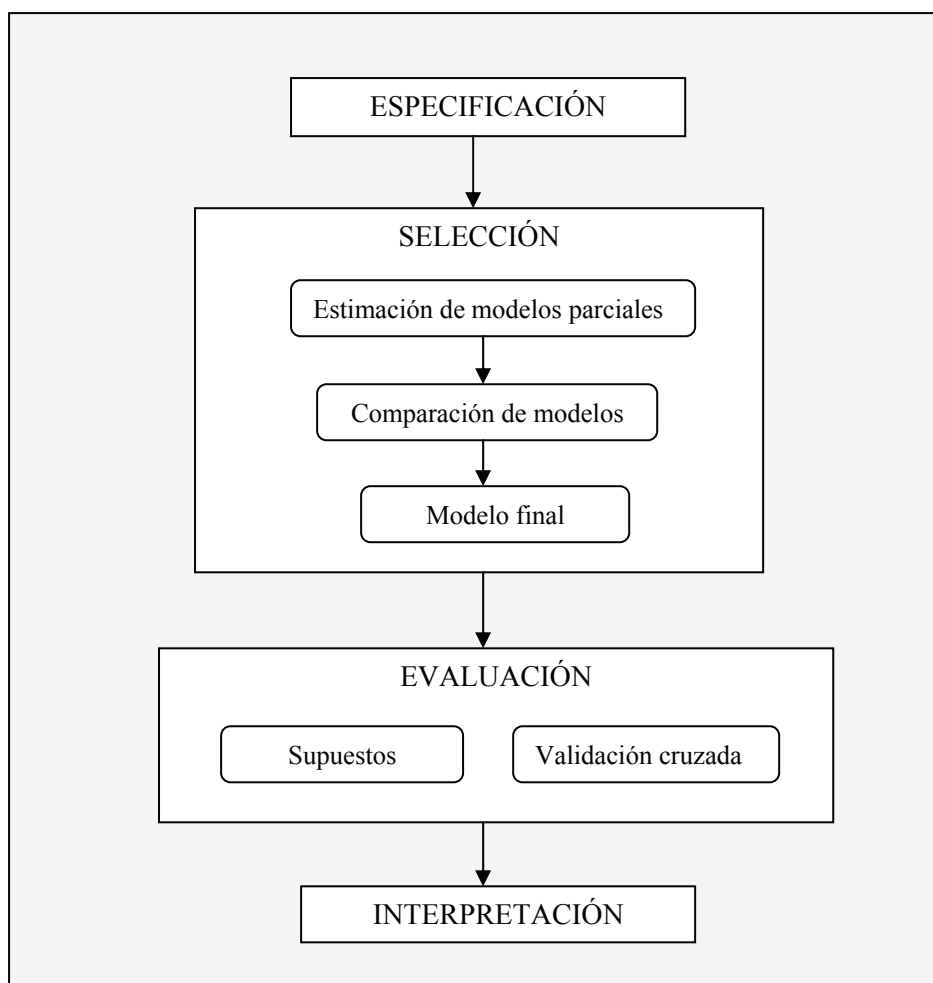


Figura 9. Proceso de modelado estadístico de los datos

3. 4. 4. 2. Selección

Una vez determinadas las variables con una relación estadísticamente significativa con la variable respuesta, se procedió a la estimación del modelo trabajando en primer término por bloques. Se comenzó estimando un modelo con las variables procedentes del ámbito cognitivo (primero las relativas al conocimiento y después otro con las correspondientes a los componentes de procesamiento), a

continuación otro modelo con las variables del ámbito metacognitivo y, por último, uno con el bloque de variables relativas a la tarea. Se realizó el modelado paso a paso, utilizando siempre y de forma sistemática dos métodos distintos de estimación, a fin de verificar que se obtenían los mismos resultados: el método de Wald hacia delante y el método de razón de verosimilitudes hacia atrás.

Se valoraron los distintos modelos obtenidos examinando:

1. la contribución de cada variable al modelo mediante el estadístico de Wald,
2. el ajuste global del modelo mediante los estadísticos de Hosmer y Lemeshow (H-L) y la razón de verosimilitud (G), el coeficiente pseudo R^2 de Nagelkerke y el análisis gráfico de residuos estudentizados y valores de influencia (el análogo en la regresión logística al estadístico de influencia de Cook) y
3. la eficacia predictiva del modelo mediante el porcentaje de casos correctamente pronosticados (% CP).

Adicionalmente, se consideraron otros indicadores importantes que podrían revelar problemas de estimación, tales como los valores del error típico de estimación de los coeficientes de regresión y la diferencia entre el valor obtenido al final del proceso y el obtenido en el modelo univariado estimado en la fase de especificación para estos coeficientes.

Con las variables obtenidas en el modelo resultante para cada bloque se estimó el modelo final, trabajando similarmente. Primeramente, hacia delante, incorporando progresivamente las variables de cada bloque en el mismo orden ya señalado, una a una, dejando en cada paso las que mostraban aportes significativos al modelo y sacando las que no. Luego hacia atrás, también paso a paso, comenzando con todas las variables seleccionadas y eliminando las que no hacían un aporte significativo al modelo. Se seleccionó el modelo cuyas variables predictoras resultaron significativas, siempre que

sus indicadores –tanto de ajuste global como de eficacia predictiva- presentaran valores apropiados.

3. 4. 4. 3. Evaluación

Para la fase de evaluación se examinó en el modelo final el cumplimiento del supuesto de linealidad entre el logit y cada predictor mediante un análisis gráfico, el supuesto de ausencia de multicolinealidad examinando la matriz de correlaciones y, para el caso del formato EA, el supuesto de igualdad de las pendientes de las funciones de regresión para todas las categorías de la variable, llevando a cabo el test de las líneas paralelas. En cuanto a la independencia de las observaciones, el cumplimiento de este supuesto se había verificado en una fase anterior, al estudiar la posible estructura jerárquica de los datos.

Seguidamente, se procedió a realizar una validación cruzada del modelo finalmente obtenido, dividiendo aleatoriamente la muestra en dos y utilizando una mitad como muestra de estimación y la otra como muestra de validación.

CAPÍTULO IV

RESULTADOS

Primero habrá que mostrar, luego, si se puede, demostrar.

Federico Fellini

4. 1. CONTROL DE CALIDAD

Se realizó un exhaustivo análisis preliminar a objeto de garantizar la mayor credibilidad y confianza en los datos reunidos, toda vez que la recogida de los mismos no pudo realizarse conforme a la planificación pautada. Además, este examen permitió un conocimiento importante sobre el comportamiento de los datos, muy útil en la comprensión de resultados posteriores.

Previo a proceder al estudio de la distribución de frecuencias de cada variable, se realizó un análisis de las respuestas emitidas por los sujetos en los 4 institutos donde no fue posible hacer la recogida de datos de modo simultáneo, a fin de garantizar que no se hubieran producido problemas de comunicación del contenido de los ítems o de las posibles soluciones, debido al hecho de no poder preservar la absoluta confidencialidad de la prueba antes de todas las aplicaciones.

En el examen de la Tabla 10 se observa que el porcentaje de respuestas correctas o sin respuesta varía entre los distintos grupos y momentos de la aplicación. No obstante, en el instituto Monay el porcentaje de respuestas correctas a cada ítem en la segunda jornada de aplicación es sospechosamente elevado, de la misma forma que el porcentaje de no respuesta también es bastante bajo en la mayoría de esos casos. Este comportamiento anómalo y el hecho de que ninguna de estas aplicaciones fue supervisada por la doctoranda condujeron a la decisión de depurar la muestra y suprimir los grupos A, B y G, sobre los cuales no se tiene suficiente confianza. En el resto de institutos, aunque para algunos grupos se observen algunos valores superiores en el porcentaje de respuestas correctas o inferiores en el porcentaje de no respuesta, no se presenta un patrón que resulte revelador de irregularidades. Sumado a esto, el hecho de que el momento de aplicación fuera inmediatamente después de la anterior o en un turno

de clase diferente en los cuales los grupos no coinciden, además de la supervisión de la doctoranda, se consideraron como evidencia suficiente de que no hubo comunicación posible.

Tabla 10. *Porcentaje de respuestas correctas y no respondidas en cada ítem para cada grupo por instituto*

Instituto	Momento de la aplicación	Grupo	EM		RC		EA	
			%C	%NC	%C	%NC	%C	%NC
La Paz	0	A	60.0	3.3	33.3	23.3	73.3	0.0
	1	B	75.7	6.1	63.6	9.1	100.0	0.0
	2	C	68.0	0.0	8.0	8.0	96.0	0.0
Coromoto	0	A	88.0	0.0	68.8	3.1	100	0.0
	1	B	88.0	0.0	81.3	3.1	78.1	6.3
LBCM	1	B	77.3	13.6	9.1	68.2	72.8	4.5
	0	J	47.4	5.3	10.5	0.0	47.4	5.3
	0	K	96.3	0.0	81.5	0.0	66.6	5.3
	1	N	50.0	3.8	57.7	7.7	42.3	15.4
	0	P	96.3	0.0	88.9	3.7	74.0	0.0
Monay	1	A	85.7	9.5	90.5	0.0	85.7	4.8
	1	B	72.0	12.0	88.0	8.0	28.0	3.2
	0	C	53.8	27.0	50.0	7.7	73.0	0.0
	0	D	44.0	20.0	52.0	12.0	64.0	4.0
	0	F	30.4	3.1	54.8	3.3	45.2	6.5
	1	G	67.6	0.0	82.4	0.0	85.3	0.0

4. 1. 1. Estudio univariado

Con este análisis se buscó conocer la distribución de las respuestas obtenidas, a fin de detectar datos fuera de rango o incongruencias que revelasen errores o inconsistencias en la base de datos y, adicionalmente, tener una primera impresión sobre la información recogida.

El 96,6% de los examinados respondió todos los ítems de la prueba general. El hecho de que un estudiante no respondiera a algún ítem era porque, en la mayoría de los casos, tampoco había respondido el correspondiente ítem en la prueba de competencias y en consecuencia parecía no sentirse capaz de valorar el nivel de dificultad de cada

componente necesario para resolver ese ítem (justamente son los ítems relativos a los componentes de procesamiento los que han sido dejados de responder).

En las Tablas 11, 12 y 13 se presenta la distribución de frecuencias de las respuestas a los ítems de la prueba de competencias. En el examen de estas tablas no se advierten anomalías que puedan indicar problemas en los datos. Sin embargo, se observa en la Tabla 12 un porcentaje llamativo de no respuesta para el ítem RC, asunto que se abordará detalladamente en el epígrafe 4. 1. 2. Por otra parte, también se observa en la tabla de frecuencias para las respuestas al ítem de elección múltiple un número elevado de examinados que eligieron la opción c a la vez que un porcentaje muy bajo de sujetos que optaron por las opciones **a** y **d**, lo que en principio pudiera dar la impresión de un funcionamiento anómalo de este distractor.

Tabla 11. *Distribución de frecuencias para las respuestas al ítem EM*

Opción de respuesta	Frecuencia	Porcentaje
a	31	6.5
b	239	50.2
c	175	36.8
d	6	1.3
No respondió	25	5.3
Total	476	100

Tabla 12. *Distribución de frecuencias para las respuestas al ítem RC*

Respuesta	Frecuencia	Porcentaje
Incorrecta	178	37.4
Correcta	244	51.3
No respondió	54	11.3
Total	476	100

Tabla 13. *Distribución de frecuencias para las respuestas al ítem EA*

Respuesta	Frecuencia	Porcentaje
Incorrecta	116	24.4
Parcialmente correcta	110	23.1
Correcta	231	48.5
No respondió	19	4.0
Total	476	100

Al estudiar el contenido del ítem de EM con detalle surge la sospecha de que los sujetos seleccionaron la respuesta basados en la observación del mapa, lo que podría explicar este alto valor de frecuencia para esa alternativa de respuesta. Para hacer una comprobación empírica de esta posibilidad se utilizó una tabla cruzada entre las respuestas al ítem de EM y el componente de procesamiento directamente implicado con la realización de cálculos numéricos (véase Tabla 14).

Tabla 14. *Relación entre las respuestas al ítem EM y la dificultad encontrada en el componente de procesamiento “hacer los cálculos”*

Opción de respuesta	Hacer los cálculos en EM						Total
	No procede	Muy difícil	Difícil	Fácil	Muy fácil	No respondió	
a	0	9	17	5	0	0	31
b	0	1	3	69	166	0	239
c	98	4	12	31	30	0	175
d	1	2	3	0	0	0	6
No respondió	0	16	6	0	0	3	25
Total	99	32	41	105	196	3	476

Se observa en la Tabla 14 que un elevado número de los sujetos que seleccionaron la opción **c** consideraron “hacer los cálculos” como un componente de procesamiento que no procede, lo cual es un fuerte indicio de que utilizaron la inspección visual del mapa como estrategia de trabajo.

También se examinó si el orden de presentación influía en la selección de esta estrategia de trabajo y si el factor tiempo pudiera estar influyendo en la forma de trabajo. Para ello, se estudió si los sujetos que optaron por la inspección del mapa sin hacer cálculos fueron examinados con la forma 4 de la prueba, donde este ítem era el último. De los 98 sujetos estudiados sólo hay 18 en la forma 4 y el resto se reparte en las otras tres formas (16 en la forma 1, 40 en la forma 2 y 24 en la forma 3). En la forma 2 el ítem con formato de EM aparece de segundo y de tercero el ítem con formato de

EA que fue respondido por el 95% de estos sujetos. Estos datos no apoyan la suposición de que la inspección visual del mapa fue una forma de trabajo determinada por el tiempo disponible para responder al ítem, ni por el orden de presentación dentro de la prueba.

Adicionalmente, se procedió a realizar una caracterización más detallada de estos 98 sujetos, de modo de verificar que son una porción de la muestra que no difiere de modo especial del resto de sujetos en su distribución en las variables contextuales.

Se observa en la Tabla 15 que estos 98 sujetos tienen una distribución en las variables sexo y edad que no difiere de manera importante con la de los sujetos que hicieron cálculos para responder el ítem, con la del total de la muestra, e incluso con la de los sujetos que seleccionaron la opción correcta b. Sin embargo, existen diferencias en la distribución por centros de los alumnos que utilizaron la estrategia visual y los que realizaron los cálculos. En particular, se observa que en los institutos Coromoto, Los Cedros y PJCM el porcentaje de sujetos que utilizan la estrategia visual es sensiblemente menor, mientras que en el Monay es considerablemente superior.

Tabla 15. *Distribución por sexo, edad y centro escolar de los sujetos que utilizan la estrategia visual en el ítem EM*

Variable	Categoría	Inspección visual	Cálculo	Total	Opción b
Sexo	Masculino	42.9	49.5	48.1	49.4
	Femenino	57.1	50.5	51.9	50.6
Edad	<15	8.2	9.3	10.1	11.8
	15	74.5	75.9	75.6	76.2
	>15	17.3	13.5	14.3	12.0
Instituto	La Paz	11.2	20.4	18.5	20.5
	MVVM	16.3	10.3	11.6	10.5
	Coromoto	5.1	15.6	13.4	18.8
	Los Cedros	3.1	7.1	6.3	8.8
	PJCM	5.1	8.2	7.6	5.4
	LBCM	26.5	25.1	25.4	27.2
	Monay	32.7	13.2	17.2	8.8

También se observó un desempeño menor de aquellos sujetos que optaron por no hacer los cálculos, reflejado en un porcentaje menor de respuestas correctas (en RC un 30.6 % frente a un 56.6 % de aciertos entre los que optaron por la estrategia de cálculo y en EA un 32.7% frente a un 52.6%) y un porcentaje importante de sujetos que no lograron contestar ningún ítem correctamente (un 34.7% frente a un 11.4% entre los que hicieron los cálculos). Esto se confirma al comparar el nivel de conocimiento declarativo, procedimental y la calificación de Matemáticas de los sujetos que utilizan como estrategia de respuesta la inspección visual y los que utilizan el cálculo, estos últimos con promedios superiores en las tres variables mencionadas.

Se comparó también el comportamiento de los sujetos que utilizaron estos dos tipos de estrategias en el conjunto de variables no cognitivas, hallándose que representan un porcentaje inferior en el uso de actividades de regulación metacognitiva y que se perciben con menor auto-eficacia, valoran peor la tarea y perciben mayor dificultad en la misma que los estudiantes que optaron por hacer los cálculos e incluso que el total de la muestra.

Lo que se deseaba verificar con este análisis era si la selección del distractor era producto del empleo de una estrategia legítima y en consecuencia se podía considerar como una respuesta parcialmente correcta. Sin embargo, a la luz de estos análisis se puede afirmar que los sujetos que optaron por hacer una inspección visual como estrategia de respuesta son menos aptos que quienes realizaron los cálculos, no sólo con un menor nivel en cuanto a conocimientos y un número menor de ítems respondidos correctamente, sino también con una valoración inferior en las variables no cognitivas consideradas en este estudio. En definitiva, optar por la inspección visual no

representa una estrategia adecuada, por lo tanto la opción **c** se considera una respuesta incorrecta, tal y como se estableció previamente.

4. 1. 2. La no respuesta

En el apartado anterior ya se comentó la alta tasa de respuesta a las preguntas de la prueba general, siendo también ésta la tónica en las preguntas de carácter contextual, en las que sólo hay 9 casos sin respuesta en la variable edad. Sin embargo, como también se indicó en el apartado anterior, la tasa de no respuesta es más alta en la prueba de competencias que en la prueba general y para el ítem con formato de respuesta corta es más del doble que para los otros dos ítems (un 11.3% frente al 4 o 5.3%).

El 83% de los estudiantes responden a los tres ítems de la prueba de competencias. Hay más varones que mujeres entre los que no responden, pero no hay diferencias sustanciales en el resto de las variables de clasificación.

La mayor tasa observada de no respuesta en la prueba de competencias parece apuntar a un problema de falta de tiempo para resolver las tres tareas propuestas. Así, se observa que el 47% de los estudiantes que no responden al ítem de EM corresponde a la forma de la prueba donde este ítem ocupaba el último lugar; el 55% de los estudiantes que no responden al ítem de RC corresponde también a la forma de la prueba donde este ítem ocupaba el último lugar y el 86% de los que dejan sin responder el ítem de EA corresponde a alguna de las dos formas donde este ítem es el último. El hecho de que en este ítem haya un porcentaje mayor de sujetos que en EM y RC se explica porque el enunciado/escenario de trabajo es completamente nuevo, algo que no sucede en los otros dos cuadernillos.

A la vista de los datos anteriores, parece fuera de toda duda que el factor velocidad afecta prácticamente por igual al ítem de EM y de RC. ¿Qué es entonces lo que explica la mayor tasa de no respuesta con el formato RC?

Para dar respuesta a esta pregunta se decidió examinar con detenimiento las características de los estudiantes que no respondían a esta pregunta y que no lo hacían por un problema de falta de tiempo (23 estudiantes que respondieron a las formas 1, 2 ó 4 de la prueba de competencias, G23). Para ello, se comparó a estos estudiantes con:

1. los estudiantes que no respondían porque les podía faltar tiempo (31 estudiantes a los que se administró la forma 3, G31).
2. el total de participantes en el estudio (GT).

La Tabla 16 proporciona la media y la desviación típica de los 3 grupos en las variables relativas al conocimiento y evidencia con claridad que los estudiantes que dejan sin responder esta pregunta y no lo hacen porque les haya faltado tiempo constituyen un grupo con un nivel de conocimientos significativamente inferior al de los otros dos grupos.

Tabla 16. *Descriptivos para las variables relativas al conocimiento*

Variable	Promedio (Desviación típica)		
	G23	G31	GT
Conocimiento declarativo	3.96 (2.163)	5.42 (2.126)	6.40 (2.125)
Conocimiento procedimental	3.96 (2.345)	5.52 (2.593)	6.48 (2.257)
Calificación académica	8.78 (5.418)	11.81 (5.256)	13.52 (4.910)

Si se analiza la actuación de estos 23 estudiantes en los otros dos ítems de la prueba de competencias esta impresión se refuerza. De estos 23 examinados solamente 1 respondió correctamente a los dos restantes ítems. El ítem de EM es respondido de forma incorrecta por 12 de estos examinados y 7 no lo responden, solo 4 lo respondieron correctamente. De la misma manera, aunque solamente 2 examinados lo

dejan sin respuesta, únicamente 4 estudiantes de este grupo de examinados logró responder correctamente al ítem de EA.

En la Tabla 17 se observa con suficiente claridad que, en comparación con el total, los sujetos que no respondieron al ítem, sean del G23 o del G31, encontraron mucha dificultad en los componentes de procesamiento cognitivo necesarios para la ejecución. Esto, sin duda, es una razón de peso para no responder el ítem.

Tabla 17. *Porcentaje de sujetos que consideraron fácil los componentes de procesamiento*

Ítem	G23	G31	GT
Comprender con claridad lo que pide el problema	39.1	29	79.9
Identificar las variables implicadas en la solución	0.0	3.2	74.5
Identificar las condiciones que deben cumplirse	8.7	3.2	78.6
Organizar la información proporcionada	0.0	0.0	73.0
Hacer cálculos	4.3	6.4	62.4
Comprobar la adecuación de la solución	4	3.2	77.1
Seleccionar o escribir la respuesta correcta	0	0	85.3

Por último, se decidió estudiar también su forma de abordar la tarea y su percepción sobre la misma. Del estudio de la Tabla 18 se deduce que aun cuando parece ser que intentaron resolver el ítem (más de la mitad afirma haber reformulado el enunciado, haber chequeado su trabajo, haber buscado una nueva manera de resolver el ítem o trabajar guiado por el análisis lógico del problema) no hacen un uso importante ni de actividades de regulación cognitiva ni de los heurísticos que pudieran haber contribuido a solucionar el ítem. En particular, el G23 no se asegura de haber comprendido la tarea para emprenderla, como lo hacen el G31 y el GT. Se observa claramente que un porcentaje muy alto encuentra dificultades para realizar la tarea, la perciben difícil, el formato les parece exigente, requiere más esfuerzo y les gusta mucho menos la tarea que al total.

Tabla 18. *Porcentaje de respuestas afirmativas en los ítems correspondientes a variables del ámbito no cognitivo*

Ítem	G23	G31	GT
Regulación metacognitiva			
¿ha leído detenidamente la pregunta?	69.6	67.7	87.8
¿se ha asegurado de haber entendido la tarea?	39.1	71.0	79.9
¿ha reformulado el enunciado?	52.2	41.9	56.5
¿ha buscado y entresacado la información principal?	26.1	29.0	48.4
¿ha hecho un pequeño plano de situación?	39.1	16.1	41.9
¿se ha planteado alguna estrategia concreta?	47.8	22.6	57.4
¿sabía perfectamente lo que tenía que hacer?	26.1	9.7	62.2
¿ha ido chequeando su trabajo?	65.2	19.4	77.2
¿ha encontrado alguna dificultad para realizar la tarea?	82.6	80.6	46.6
Seguir adelante como si nada	39.1	28.0	23.6
Parar y leer de nuevo el enunciado de la pregunta	43.5	64.0	70.3
Buscar una nueva manera de resolver la pregunta	52.2	52.0	70.3
Conocimiento estratégico			
¿Consideró varias formas de trabajo para responder?	13	29	45.7
¿Ha tratado de dividir el problema en pequeños pasos?	13	19.4	47.8
¿Planteó una situación similar?	21.7	12.9	40.8
¿Descartó las alternativas que no le parecían posibles?	26.1	16.1	57.0
¿Utilizó procedimientos indirectos?	8.7	3.2	35.1
¿Respondió partiendo de una respuesta probable?	13	12.9	52.7
¿Trabajó con una estrategia que ha utilizado antes?	26.1	9.7	29.1
¿Trabajó guiado por el análisis lógico del problema?	52.2	32.3	61.3
Percepción sobre la tarea			
La tarea planteada es entretenida	52.2	22.6	74.4
La pregunta permite demostrar lo que uno sabe	56.5	32.3	78.6
La tarea es de poca utilidad para mi formación	47.8	54.8	56.3
Me gusta la tarea que se plantea en la pregunta	39.1	22.6	75.4
La pregunta me pareció difícil	82.6	80.6	43.9
Para contestar bien es necesario saber mucho	56.5	74.2	38.7
Las preguntas con este formato son tediosas	52.2	61.3	36.6
Es importante saber responder a este formato	65.2	48.4	81.7
Me gustan las preguntas con este formato	30.4	25.8	71.8
Trabajar en este formato puede ser de utilidad	78.3	41.9	85.8
Con este formato es fácil aprobar un examen	43.5	25.8	78.6
Este formato exige mucho esfuerzo	65.2	58.1	30.0
En este formato la pregunta parece más difícil	65.2	58.1	26.9

Del examen de toda esta información puede decirse que estos 23 estudiantes son un grupo de nivel inferior en las variables estudiadas y que esto ha podido dar lugar a la falta de respuesta al ítem.

Sin embargo, la pregunta es ¿por qué responden las otras dos tareas? La razón podría estar en que, si bien es cierto que el ítem RC comparte el mismo enunciado del ítem EM, la respuesta no puede ser emitida al azar, como pudiera haber sido la forma de proceder para responder el ítem EM, dado que sólo 4 cuatro de ellos emitieron la respuesta correcta. Por otro lado, también es cierto que el ítem RC tiene un formato de construcción como el del ítem EA, pero el ítem RC de la prueba de competencias tiene una única solución y no permite una elaboración de la misma por tanteo como en el caso del ítem EA. Una explicación absolutamente clara escapa del análisis preliminar y seguramente podría estar en esas variables que se hipotetizan como influyentes en la ejecución de un ítem. Importante resulta el hecho de que el formato de la tarea ya marca una diferencia.

4. 1. 3. Patrones anómalos de respuesta

Con el fin de detectar posibles inconsistencias se elaboraron tablas cruzadas con distintos indicadores del mismo constructo. Los resultados muestran que las respuestas de los sujetos son consistentes, es decir, su respuesta a un ítem tiene la dirección esperada en el otro indicador sobre el mismo constructo. Por señalar sólo algunos ejemplos, los sujetos a quienes les pareció el ítem difícil están de acuerdo en que es necesario saber mucho para contestarlo, o los sujetos que están de acuerdo en creer que contestaron correctamente el ítem consideran que saben lo suficiente como para responderlo bien.

Asimismo, se examinó y descartó la presencia de patrones de respuesta aquiescente y de respuesta extrema.

4. 1. 4. Calidad psicométrica

A continuación se presentan los resultados obtenidos al examinar las propiedades psicométricas de las dos pruebas utilizadas en el estudio. En ambos casos se ha examinado la calidad métrica de la prueba considerada en su conjunto. En el caso de la prueba de competencias, se ha examinado también la calidad métrica de cada una de las 3 preguntas de la prueba, dado el papel clave que desempeñan en la investigación.

4. 1. 4. 1. Prueba de competencias

En las Figuras 10, 11 y 12 se presenta la curva característica y la curva de información de los 3 ítems de la prueba de competencias, mostrándose en la Tabla 19 los valores obtenidos para los parámetros de estos ítems, obtenidos tanto desde el marco de la teoría de respuesta al ítem como desde la teoría clásica.

Los tres ítems presentan una dificultad semejante y una buena capacidad discriminativa, ligeramente inferior en el ítem de evaluación de la actuación, con una precisión máxima en la zona central de la escala y una correlación bastante alta con la calificación en Matemáticas.

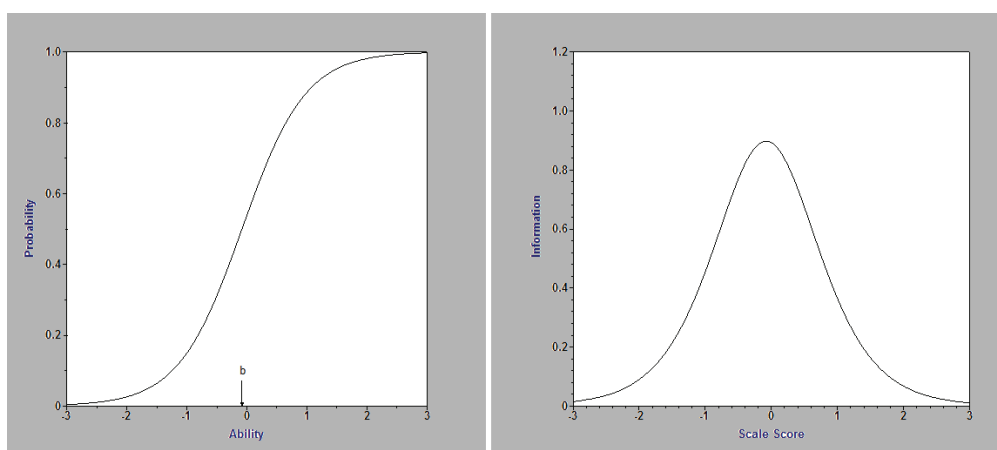


Figura 10. Curva característica y función de información del ítem EM

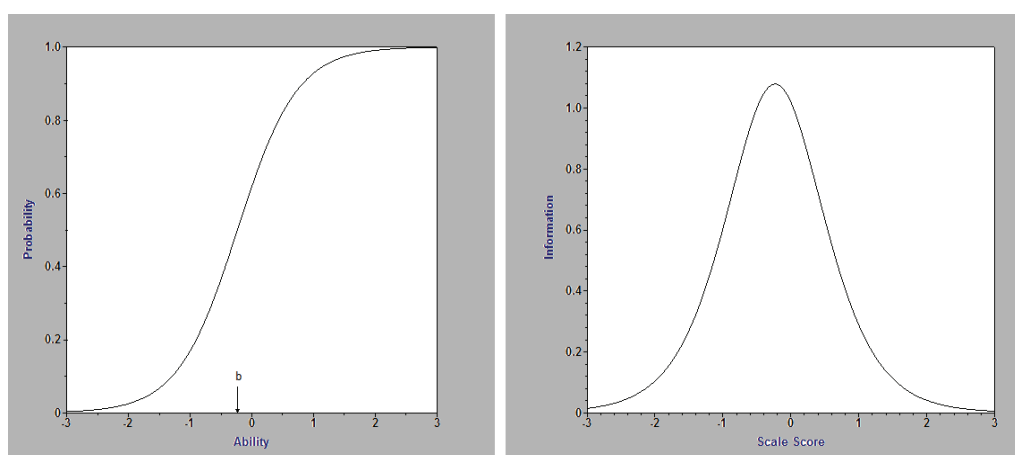


Figura 11. Curva característica y función de información del ítem RC

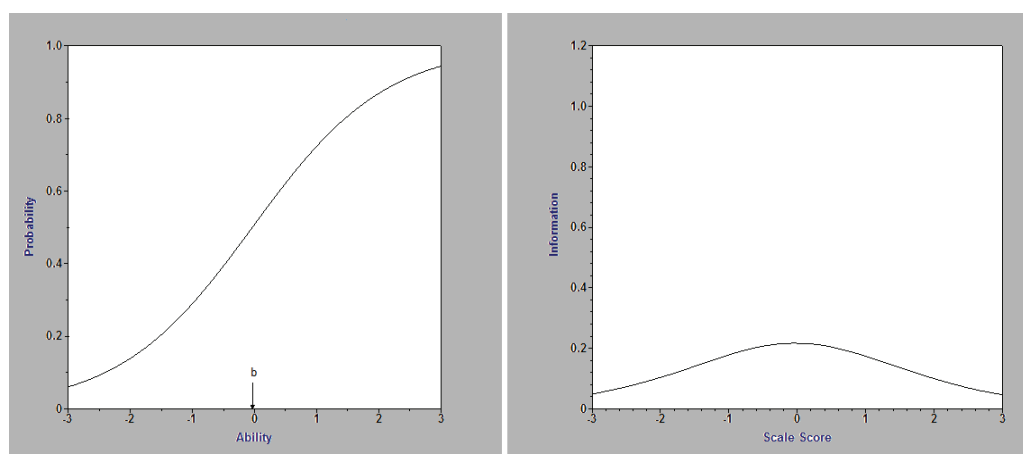


Figura 12. Curva característica y función de información del ítem EA

Tabla 19. *Parámetros e índices de fiabilidad y validez de los ítems de la prueba de competencias*

	EM	RC	EA
Discriminación (TCT)	0.380	0.393	0.305
Discriminación (TRI)	1.895	2.077	0.930
Dificultad (TCT)	0.530	0.578	0.505
Dificultad (TRI)	-0.081	-0.233	-0.032
Índice de fiabilidad	0.190	0.194	0.255
Índice de validez	0.224	0.240	0.464
Correlación con CM	0.447	0.486	0.556

El valor obtenido para el coeficiente alfa de Cronbach en la prueba es de 0.519 y la curva de información del test se muestra en la Figura 13.

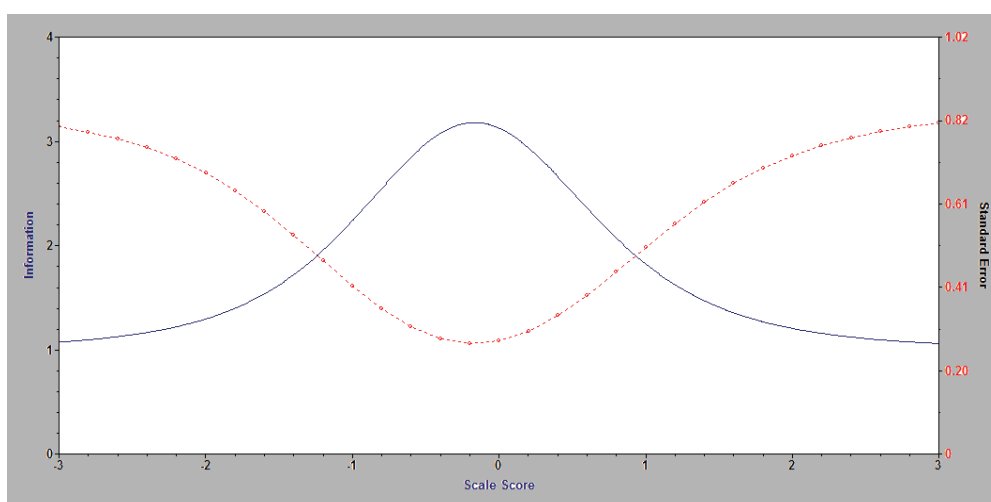


Figura 13. Función de información del test

El análisis factorial exploratorio revela la presencia de un único factor, en el que cargan los tres ítems con pesos factoriales altos y bastante similares: 0.751 para EM, 0.792 para RC y .691 para EA. Por otro lado, la capacidad de la prueba de competencias para pronosticar el rendimiento académico en Matemáticas es verdaderamente alta, habida cuenta además del reducido número de ítems: 0.70 es el valor de la correlación entre la puntuación total en la prueba de los estudiantes y la última calificación en dicha asignatura proporcionada por su profesor.

En resumen, las evidencias recogidas tanto acerca de la precisión como de la validez de las puntuaciones obtenidas con los ítems de esta prueba apoyan su utilización para dar respuesta a la pregunta de investigación planteada en este estudio.

4. 1. 4. 2. Prueba general

En la Tabla 20 se informa el número de ítems y los valores del coeficiente alfa para cada grupo de ítems según variable a medir. Los valores observados permiten afirmar que las puntuaciones obtenidas con los ítems de la prueba general son fiables y que están en perfecta consonancia con los reportados para las escalas revisadas que

usualmente se utilizan para la medición de estas variables (valores de alfa que van de 0.62 a 0.93).

Tabla 20. *Fiabilidad de los subconjuntos de ítems por variable*

Variable	N° de ítems	EM	RC	EA
CAP	3	0.749	0.782	0.786
VTC	4	0.708	0.733	0.709
VTF	4	0.607	0.681	0.691
VPT	8	0.766	0.809	0.809
DTC	2	0.609	0.603	0.622
DTF	3	0.547	0.639	0.653
DPT	5	0.619	0.715	0.707
COD	3	0.771	0.765	0.789
COM	3	0.622	0.900	0.867
CPC	7	0.827	0.929	0.918

Los resultados de los análisis factoriales confirmatorios realizados permiten corroborar la estructura interna hipotetizada para cada variable medida. Las Tablas 21 a 24 muestran los valores obtenidos en los estadísticos indicadores del ajuste para cada uno de los modelos planteados.

Tabla 21. *Índices para el diagnóstico del AFC para valor percibido de la tarea*

Índices	EM	RC	EA
RMSEA	0.039-0.078	0.037-0.077	0.046-0.084
GFI	0.991	0.994	0.992
AGFI	0.984	0.988	0.986
Chi cuadrado (p)	38.18 (0.006)	42.30 (0.002)	57.62 (0.0001)

Tabla 22. *Índices para el diagnóstico del AFC para dificultad percibida de la tarea*

Índices	EM	RC	EA
RMSEA	0.056-0.135	0.008-0.100	0.00-0.077
GFI	0.985	0.997	0.998
AGFI	0.945	0.988	0.992
Chi cuadrado (p)	20.57 (0.001)	9.79 (0.044)	5.05 (0.282)

Tabla 23. *Índices para el diagnóstico del AFC para metacognición*

Índices	EM	RC	EA
RMSEA	0.030-0.142	0.00-0.126	0.00-0.104
GFI	0.995	0.998	0.999
AGFI	0.976	0.994	0.996
Chi cuadrado (p)	8.31 (0.016)	5.78 (0.055)	3.17 (0.205)

Tabla 24. *Índices para el diagnóstico del AFC para componentes de procesamiento*

Índices	EM	RC	EA
RMSEA	0.00-0.057	0.00-0.000	0.026-0.080
GFI	0.995	1.000	0.998
AGFI	0.987	0.999	0.995
Chi cuadrado (p)	13.77 (0.246)	3.59 (0.980)	17.03 (0.107)

En resumen, los resultados obtenidos para el coeficiente alfa y para los análisis confirmatorios representan evidencias favorables acerca de la fiabilidad de las puntuaciones obtenidas con la prueba general y de la validez de las inferencias a que pueda dar lugar.

4. 1. 5. Información recabada a los profesores

Con el objeto de verificar si las valoraciones proporcionadas por los docentes de matemáticas sobre el nivel de conocimiento declarativo y procedimental de sus alumnos están guiadas más por la comprensión de lo que se les señalara en la definición de cada variable relativa al tipo de conocimiento, que por una simple relación con la calificación en la asignatura, se procedió a realizar un estudio de las correlaciones entre las valoraciones de estos dos tipos de conocimiento en los grupos de rendimiento bajo, medio y alto de sujetos, usando como criterio de clasificación tanto la calificación en matemáticas como el desempeño en la prueba de competencias (véase Tabla 25). Los distintos valores obtenidos para el coeficiente de correlación en los 3 grupos examinados, así como la consistencia en el patrón descendente de los mismos cuando los grupos se forman en base a la calificación del profesor y al rendimiento de los alumnos en la prueba de competencias pone de manifiesto que los profesores realizaron sus valoraciones del nivel de conocimiento declarativo y procedimental de los alumnos

siguiendo las indicaciones proporcionadas y no guiados por el valor de la calificación académica.

Tabla 25. *Valores de correlación entre conocimiento declarativo y procedimental*

GRUPO	VARIABLE PARA FORMAR LOS GRUPOS	
	Calificación en Matemáticas	Desempeño en la prueba de competencias
Inferior	0.843	0.916
Central	0.637	0.810
Superior	0.584	0.684
Total	0.914	0.914

En términos generales, con base en todos los análisis preliminares realizados a la muestra de sujetos examinados y en relación a la calidad psicométrica de las puntuaciones obtenidas se puede afirmar que los datos reunidos poseen la coherencia y credibilidad suficientes y necesarias para certificar el valor de los resultados obtenidos y la validez de las conclusiones a extraer del análisis de los mismos.

4. 2. ESTRUCTURA DE LOS DATOS

Con el fin de determinar si existe una estructura jerárquica en los datos se realizó un análisis de varianza de un factor de efectos aleatorios.

En la Tabla 26 se presentan los valores del coeficiente de correlación intra-clase (CCI), que indican que la variabilidad entre centros representa un porcentaje bajo de la variabilidad total para cualquiera de los formatos. También se observa que la probabilidad asociada a la prueba de Wald para cualquiera de los casos es superior a 0.05. Por tanto, se acepta la hipótesis nula que establece que la varianza del factor es cero y no existe efecto del instituto, por lo que no es necesario utilizar una aproximación multinivel al modelar los datos.

Tabla 26. *Análisis de varianza de los institutos*

Variable respuesta	Varianza del factor	Varianza de los residuos	CCI	Z de Wald (p)
Respuesta EM	0.025	0.690	0.035	1.395 (0.163)
Respuesta RC	0.021	0.229	0.083	1.395 (0.163)
Respuesta EA	0.066	0.640	0.094	1.489 (0.137)
Nº de resp. correctas	0.089	1.120	0.0735	1.428 (0.153)

Adicionalmente, en razón de la forma en que debió hacerse la aplicación, se realizó este análisis anidando los datos en función de la forma de la prueba. En la Tabla 27 se presentan los resultados del análisis de varianza en la que se observan valores de CCI no muy elevados, aunque algo mayores que los relativos a los institutos. No obstante, la Z de Wald en la variable respuesta para cualquiera de los formatos tiene significación superior a 0.05, lo cual nos lleva a la aceptación de la hipótesis nula que establece que la varianza del factor es cero y en consecuencia que no existe un efecto atribuible a la forma de la prueba.

Tabla 27. *Análisis de varianza de la forma de la prueba de competencias*

Variable respuesta	Varianza del factor	Varianza de los residuos	CCI	Z de Wald (p)
Respuesta EM	0.053	0.217	0.196	1.171 (0.242)
Respuesta RC	0.006	0.241	0.023	0.838 (0.402)
Respuesta EA	0.100	0.635	0.136	1.143 (0.253)

Con base en estos resultados se puede afirmar que ni el instituto, ni la forma de la prueba que le correspondió a cada sujeto examinado producen un efecto tan acentuado sobre la respuesta al ítem que haga imprescindible un análisis multinivel.

4. 3. EXAMEN PRELIMINAR DE LAS VARIABLES PREDICTORAS

Una vez construidas las variables predictoras del modelo según se indicó en el

epígrafe 3. 4. 3. del capítulo correspondiente a Método, se calcularon los estadísticos descriptivos para las mismas, cuyos valores se muestran en la Tabla 28.

Tabla 28. *Estadísticos descriptivos de las variables predictoras*

Variable	Media (Desviación típica)			Rango
	EM	RC	EA	
CODI	9.10 (2.215)	8.10 (2.466)	8.36 (2.478)	3-12
COMB	8.66 (2.701)	7.60 (2.906)	8.07 (2.801)	3-12
RESP	3.43 (0.850)	2.57 (1.120)	2.61 (1.006)	1-4
CAP	9.10 (2.245)	8.42 (2.641)	8.67 (2.875)	3-12
PLAN	2.38 (0.795)	2.43 (0.811)	2.50 (0.809)	0-3
CONT	2.83 (1.234)	2.83 (1.268)	3.11 (1.111)	0-4
CES	3.29 (1.634)	3.18 (1.790)	3.02 (1.718)	0-7
VTC	12.14 (3.220)	11.69 (3.326)	12.10 (3.323)	4-16
VTF	12.53 (2.781)	11.88 (3.068)	12.22 (3.110)	4-16
DTC	4.46 (1.902)	4.66 (1.953)	4.85 (2.014)	2-8
DTF	5.76 (2.225)	6.77 (2.467)	7.01 (2.643)	3-12
CD		6.39 (2.125)		0-10
CP		6.49 (2.257)		0-10
CM		13.43 (4.910)		0-20

El examen de la Tabla 28 da algunos indicios del efecto del formato. Es bastante llamativo como se diferencian los valores de las distintas variables en los formatos EM y RC, siendo que ambos ítems implican la misma tarea. Se muestra como el ítem EM es valorado como el más fácil en todas las fases de procesamiento cognitivo, que a su vez son consideradas más difíciles para el ítem RC, evidenciando que la forma cómo resuelven los ítems de EM y RC es bien diferente (véase también la Tabla 29).

Tabla 29. *Porcentaje de examinados que encuentran fáciles los componentes de procesamiento para los formatos EM y RC*

Ítem	EM	RC
Comprender con claridad lo que pide el problema	79.0	73.9
Identificar las variables implicadas en la solución	74.6	52.1
Identificar las condiciones que debe cumplir la solución	77.7	53.6
Organizar la información proporcionada	72.9	50.2
Hacer cálculos	63.2	54.8
Comprobar la adecuación de la solución	77.5	52.1
Seleccionar o escribir la respuesta correcta	85.7	52.3

Continuando con el análisis de los resultados presentados en la Tabla 28, se observa que, en relación a las variables del ámbito metacognitivo, los examinados se consideran más competentes en la tarea en formato EM y menos en la misma tarea en formato RC.

En cambio, para las actividades de control ejecutivo (planificación y control) se observan valores muy similares, lo que podría ser un indicador de que el uso de la regulación metacognitiva es independiente de la forma de la tarea.

En cuanto al conocimiento estratégico, el número de estrategias cognitivas utilizadas según formato es aproximadamente el mismo. Pero particularmente en esta variable es importante señalar el alcance de este resultado y tratar de verificar si el uso de cada estrategia es adecuado, pues podría ocurrir que el empleo de un heurístico no siempre lleve a la respuesta correcta.

Con este objetivo, se construyeron tablas cruzadas entre la respuesta al ítem y la estrategia utilizada. En la Tabla 30 se observa que no existe relación entre la respuesta al ítem EM y estrategias como plantear situaciones similares, emplear una estrategia utilizada en ocasiones anteriores y guiarse por el análisis lógico del problema. Además, de las cuatro estrategias donde existe relación, solo 2 parecen conducir a la solución del problema (dividir el problema en pequeños pasos y utilizar procedimientos indirectos). Para los formatos RC y EA se observa una relación significativa con todas las estrategias propuestas resultando todas estas efectivas, es decir, con un número mayor de respuestas correctas cuando se utiliza la estrategia que cuando no (véase Tabla 31).

Por su parte, los resultados presentados en la Tabla 28 sobre las variables relacionadas a la percepción sobre la tarea muestran que aun cuando los valores sobre valor percibido para las tres tareas son muy similares la valoración más alta corresponde

al ítem de EM, tanto en relación al contenido como al formato y la peor para el ítem de RC, también en los dos aspectos y pese a que el contenido es idéntico al del formato de EM. En cuanto a la dificultad percibida de la tarea, cuando se trata de considerarla según el contenido de la tarea la percepción es más similar, mientras que cuando se trata del formato utilizado para demandarla, la tarea planteada en formato de EM es considerada la más fácil. Todo esto supone una clara influencia del formato por cuanto los ítems de EM y RC comparten el enunciado y a pesar de ello no son consideradas de la misma manera.

Tabla 30. *Relación entre las estrategias utilizadas al responder a los ítems de la prueba de competencias y la actuación en cada uno de ellos*

Estrategia	χ^2 (p)		
	EM	RC	EA
Trató de dividir en pasos	10.159 (0.006)	128.499 (0.000)	110.573 (0.000)
Planteó otra situación similar	0.366 (0.833)	88.765 (0.000)	86.478 (0.000)
Descartó alternativas	26.766 (0.000)	68.513 (0.000)	19.194 (0.000)
Utilizó modos indirectos	12.901 (0.002)	60.782 (0.000)	22.874 (0.000)
Partió de una respuesta probable	70.057 (0.000)	50.389 (0.000)	9.736 (0.021)
Empleo una estrategia ya usada	4.576 (0.101)	7.174 (0.028)	16.217 (0.001)
Se guio por el análisis lógico	1.528 (0.466)	6.876 (0.032)	65.105 (0.000)

Tabla 31. *Relación entre la respuesta al ítem según formato y la estrategia utilizada*

Estrategia	EM		RC		EA	
	0	1	0	1	0	2
Trató de dividir en pasos	90	132	45	183	25	171
Planteó otra situación similar	88	101	46	162	26	165
Descartó alternativas	144	135	62	167	20	87
Utilizó modos indirectos	60	104	43	128	35	119
Partió de una respuesta probable	157	91	69	159	16	63
Empleo una estrategia ya usada	55	81	64	76	28	91
Se guio por el análisis lógico	125	153	102	147	42	189

Adicionalmente, antes de proceder al modelado se hizo una comparación de medias para examinar la existencia de diferencias en las variables predictoras según formato. Si bien es cierto que para todos los conjuntos de variables la diferencia resultó ser significativa, es importante resaltar que el tamaño del efecto es muy pequeño.

Habría que destacar las variables relativas a los componentes de procesamiento, particularmente codificación ($\eta^2 = 0.162$) y respuesta ($\eta^2 = 0.389$), la dificultad percibida de la tarea en relación al formato ($\eta^2 = 0.202$) y el conocimiento auto-percibido ($\eta^2 = 0.149$), con valores del tamaño del efecto, que si bien no son altos, son superiores al resto, con una potencia alta.

En la Tabla 32 se pueden ver las diferencias entre las medias para las variables mencionadas entre cada par de ítems con formato diferente. En ella se observa que la diferencia se presenta en el ítem de elección, tanto con la tarea de respuesta corta (con la que comparte el mismo enunciado) como con la de evaluación de la actuación, mientras que no se presenta diferencia entre los ítem de respuesta construida. Este resultado es de notable interés porque informa de cómo el formato representa un factor que subyace como determinante de la ejecución del ítem, tanto en lo concerniente a los componentes de procesamiento como también en relación a lo no estrictamente cognitivo, tal como la percepción que tiene el sujeto tanto sobre sí mismo como sobre la dificultad de la tarea vinculada a la presencia o no del formato de elección.

Tabla 32. *Diferencias de medias entre formatos por variable*

Variable	Diferencia entre medias (p)		
	EM RC	EM EA	RC EA
CODI	1.023 (0.000)	0.760 (0.000)	0.264 (0.133)
RESP	0.861 (0.000)	0.812 (0.000)	0.049 (1.000)
CAP	0.674 (0.000)	0.424 (0.008)	0.250 (0.313)
DTF	1.008 (0.000)	1.250 (0.000)	0.242 (0.207)

En esta línea de búsqueda de diferencias entre los posibles formatos del ítem se analizó el valor de los coeficientes de correlación de cada predictor y la respuesta al ítem. Los coeficientes de correlación se muestran en la Tabla 33 y se observa que la mayor fuerza de asociación, en general, está en el formato de evaluación de la

actuación, seguido del formato de respuesta corta y por último, el formato de elección múltiple, que presenta la menor fuerza de asociación con las variables predictoras.

Destacable el resultado obtenido entre las variables componentes de procesamiento y la respuesta al ítem de respuesta corta, donde se presentan los valores más altos para los coeficientes de correlación.

Tabla 33. *Coefficiente de correlación entre la variable respuesta y los predictores en cada formato*

Variable	Coeficiente de correlación (p)		
	EM	RC	EA
CM	0.447 (0.000)	0.486 (0.000)	0.556 (0.000)
CD	0.533 (0.000)	0.526 (0.000)	0.563 (0.000)
CP	0.477 (0.000)	0.534 (0.000)	0.654 (0.000)
CODI	0.378 (0.000)	0.789 (0.000)	0.733 (0.000)
COMB	0.744 (0.000)	0.860 (0.000)	0.851 (0.000)
RESP	0.292 (0.000)	0.861 (0.000)	0.734 (0.000)
CAP	0.399 (0.000)	0.598 (0.000)	0.660 (0.000)
PLAN	0.099 (0.036)	0.437 (0.000)	0.546 (0.000)
CONT	0.611 (0.000)	0.690 (0.000)	0.733 (0.000)
CES	-0.017 (0.717)	0.508 (0.000)	0.511 (0.000)
VTC	0.279 (0.000)	0.303 (0.000)	0.444 (0.000)
VTF	0.346 (0.000)	0.391 (0.000)	0.525 (0.000)
DTC	-0.268 (0.000)	-0.305 (0.000)	-0.379 (0.000)
DTF	-0.288 (0.000)	-0.368 (0.000)	-0.509 (0.000)

Estos valores de correlación reflejan que las variables que tienen mayor influencia en la ejecución de la tarea son las cognitivas, particularmente los componentes de procesamiento con fuertes relaciones con las respuestas a los ítems con formato de respuesta construida, tanto si es de RC como si es de EA. También es notorio que la actividad de control metacognitivo puede resultar un importante predictor en el modelo. Otro resultado que se percibe del examen de la Tabla 33 es un orden en la importancia de cada grupo de variables para la solución del problema, que sería variables cognitivas como las más influyentes, luego las variables metacognitivas y menos influyentes las variables relacionadas con la percepción sobre la tarea.

4. 4. ESTIMACIÓN Y PRUEBA DEL MODELO

A continuación se ofrecen los resultados obtenidos al modelar las respuestas de los estudiantes a cada uno de los ítems de la prueba de competencias. La presentación de los resultados se realiza para cada uno de los formatos siguiendo el enfoque del modelado descrito en la Figura 9 del capítulo III relativo al Método. Seguidamente se comparan los modelos obtenidos para los 3 formatos examinados.

4. 4. 1. Modelo EM

4. 4. 1. 1. Especificación

De las variables teóricamente seleccionadas para el modelo, únicamente se excluyó para el análisis multivariante la variable conocimiento estratégico, al no mostrar una relación significativa con la respuesta de los alumnos en el ítem de elección múltiple ($p=0.707$).

4. 4. 1. 2. Selección

Al trabajar individualmente con los cuatro bloques de variables consideradas en el modelo, los predictores que mostraron un peso significativo fueron:

- a. Conocimiento declarativo, dentro del bloque Conocimiento
- b. Combinación, dentro del bloque Componentes de procesamiento
- c. Conocimiento autopercebido y control, dentro del bloque metacognitivo
- d. Todas las variables definidas en el bloque relativo a las tareas, el valor y la dificultad percibidas en relación tanto al formato como al contenido de la pregunta.

Se procedió seguidamente a la estimación del modelo final incorporando las variables seleccionadas de cada bloque, operando de la forma ya indicada, obteniéndose

finalmente con ambos métodos el modelo que se presenta en la Tabla 34, en el cual las variables conocimiento declarativo, combinación y control muestran que hacen un aporte estadísticamente significativo en la explicación de la respuesta al ítem EM, el resto de variables tienen valores para el estadístico de Wald con significación superior a 0.1. Los estadísticos H-L, G y R^2 de Nagelkerke muestran valores que indican que el modelo trabaja satisfactoriamente. El porcentaje de casos correctamente pronosticados revela una excelente capacidad predictiva del modelo finalmente seleccionado.

Tabla 34. Resultados del modelo de regresión logística para el formato EM

Variables	B	ET	Wald	gl	Sig.	Exp (B)	Límite inferior	Límite superior
CD	0.520	0.104	24.775	1	0.000	1.681	1.370	2.603
COMB	0.999	0.115	75.173	1	0.000	2.715	2.166	3.403
CONT	1.251	0.190	43.380	1	0.000	3.493	2.407	5.068
Constante	-16.119	1.603	101.076	1	0.000	0.000		
Ajuste global								
H-L (sig.)		G (sig.)		R^2 de Nagelkerke		% CP		
0.615 (0.961)		421.262 (0.000)		0.810		90.7		

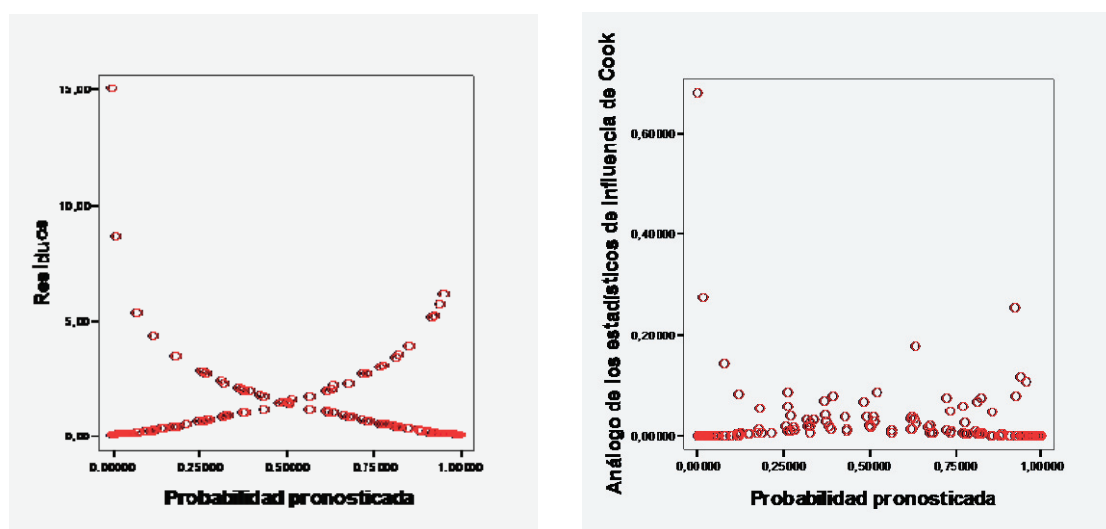


Figura 14. Gráficos de residuos y de influencia para el ítem de EM

En cuanto al estudio de los residuales y valores de influencia puede decirse que el número de casos atípicos (a más de 2 desviaciones típicas) es reducido (sólo 8 casos), lo cual se refuerza con el examen de los gráficos de la Figura 14, que muestran muy pocos casos anómalos, apoyando la bondad del ajuste de los datos al modelo.

4. 4. 1. 3. Evaluación

Para la evaluación del cumplimiento del supuesto de linealidad se elaboraron los gráficos correspondientes y los resultados se muestran en las Figuras 15, 16, 17 y 18.

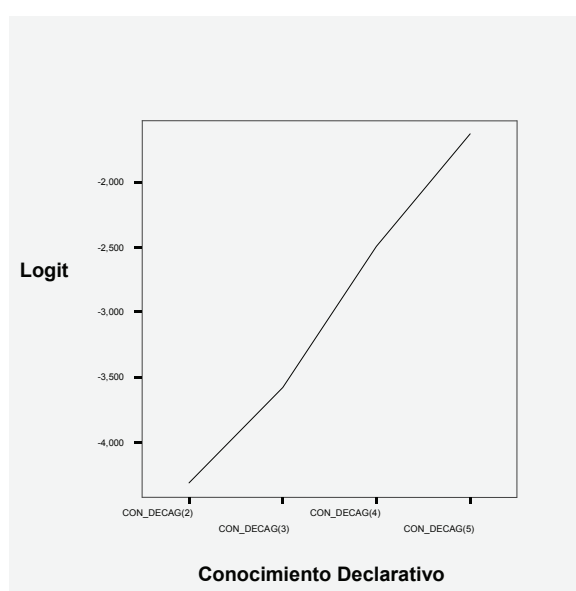


Figura 15. Relación entre el logit estimado y la variable CD para el formato EM

A la vista de los resultados obtenidos es posible afirmar que no hay una violación seria del supuesto de linealidad ni para la variable conocimiento declarativo, ni para la variable combinación. No se cumple estrictamente para la variable control. No obstante este incumplimiento, la relación entre la variable y la probabilidad pronosticada de responder correctamente es monótona creciente, tal y como se puede observar en la Figura 18. Además, la falta de linealidad no pareciera haber afectado de

modo severo las estimaciones del modelo a tal punto de esconder relaciones entre las variables predictoras y la variable dependiente; es decir, que el grado de relación no parece estar infraestimado a pesar del incumplimiento del supuesto.

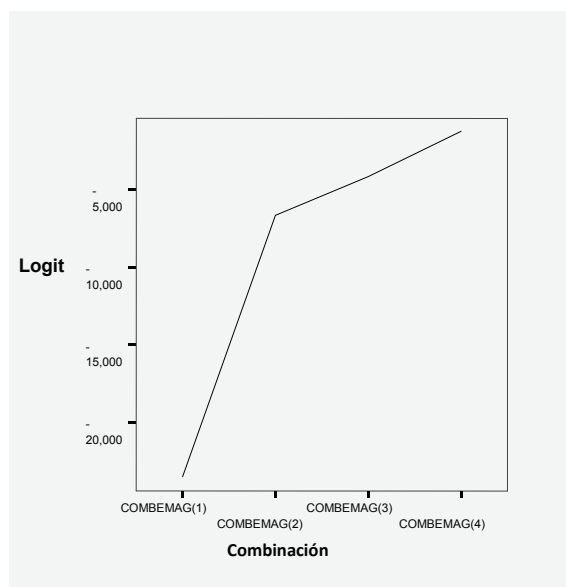


Figura 16. Relación entre el logit estimado y la variable COMB para el formato EM

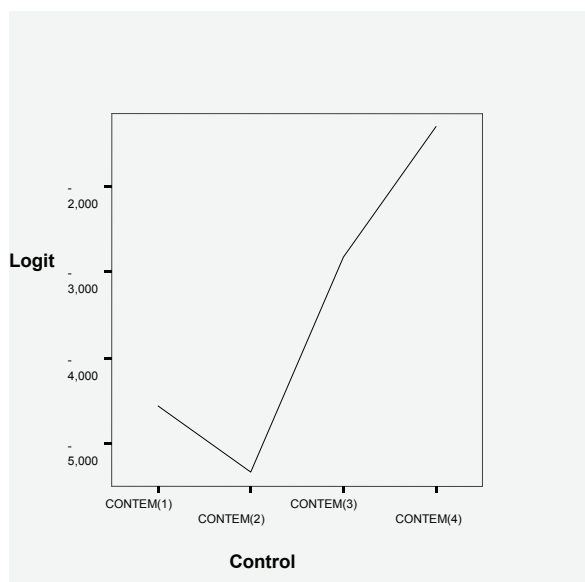


Figura 17. Relación entre el logit estimado y la variable CONT para el formato EM

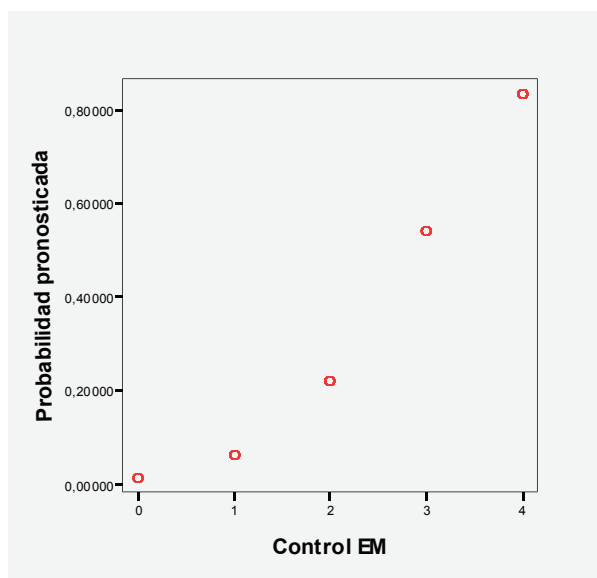


Figura 18. Relación entre la variable CONT y la probabilidad de respuesta correcta

La evaluación de la posible existencia de multicolinealidad entre las variables del modelo se hizo a través del examen de la matriz de correlaciones, la cual muestra que no existe ningún valor que pueda considerarse alarmante ($CD-COMB = 0.525$, $CD-CONT = 0.354$ y $COMB-CONT = 0.545$).

La evidencia proporcionada por la validación cruzada del modelo también es favorable: la tasa de predicciones correctas en la muestra de validación (91.2%) es incluso ligeramente superior a la obtenida al estimar el modelo con todos los sujetos de la muestra (véase Tabla 34).

4. 4. 1. 4. Interpretación

Los resultados obtenidos indican que las variables que ejercen influencia en la respuesta al ítem cuando es formulado con un formato de EM son el conocimiento declarativo, la fase de combinación de los componentes de procesamiento cognitivo y el control como actividad de regulación metacognitiva. Las variables relativas a la

percepción sobre la tarea no muestran una influencia importante cuando son consideradas conjuntamente con el resto de variables que se pusieron a prueba.

4. 4. 2. Modelo RC

4. 4. 2. 1. Especificación

Para el análisis multivariante se incorporaron todo el conjunto de variables seleccionadas en base a la teoría, dado que todas ellas mostraron relación significativa con la respuesta de los examinados en el ítem de respuesta corta.

4. 4. 2. 2. Selección

Al trabajar individualmente con los cuatro bloques de variables consideradas en el modelo, los predictores que mostraron un peso significativo fueron:

- a. Conocimiento declarativo y procedimental, dentro del bloque Conocimiento
- b. Codificación y combinación, dentro del bloque Componentes de procesamiento
- c. Todas las variables que conforman el bloque metacognitivo: conocimiento autopercebido, planificación, control y conocimiento estratégico
- d. Todas las variables que conforman el bloque relativo a las tareas: valor percibido en relación tanto al contenido como al formato y la dificultad percibida también en relación tanto al contenido como al formato de la pregunta.

Se procedió seguidamente a la estimación del modelo final incorporando las variables seleccionadas de cada bloque, operando de la forma ya indicada, obteniéndose finalmente con ambas estrategias el modelo que se presenta en la Tabla 35, en el cual

las variables codificación y combinación muestran que hacen un aporte estadísticamente significativo en la explicación de la respuesta al ítem de RC, el resto de variables tienen valores para el estadístico de Wald no significativos. Los estadísticos H-L, G y R^2 de Nagelkerke muestran valores indicativos de que el modelo trabaja satisfactoriamente. El porcentaje de casos correctamente pronosticados revela una excelente capacidad predictiva del modelo finalmente seleccionado.

Tabla 35. Resultados del modelo de regresión logística para el formato RC

Variables	B	ET	Wald	gl	Sig.	Exp (B)	Límite inferior	Límite superior
CODI	1.880	0.709	7.035	1	0.008	6.550	1.633	26.267
COMB	3.929	1.117	12.378	1	0.000	50.863	5.699	453.961
Constante	-44.724	13.122	11.616	1	0.001	0.000		
Ajuste global								
H-L (sig.)	G (sig.)		R^2 de Nagelkerke		% CP			
0.058 (1.000)	560.948 (0.000)		0.989		99.5			

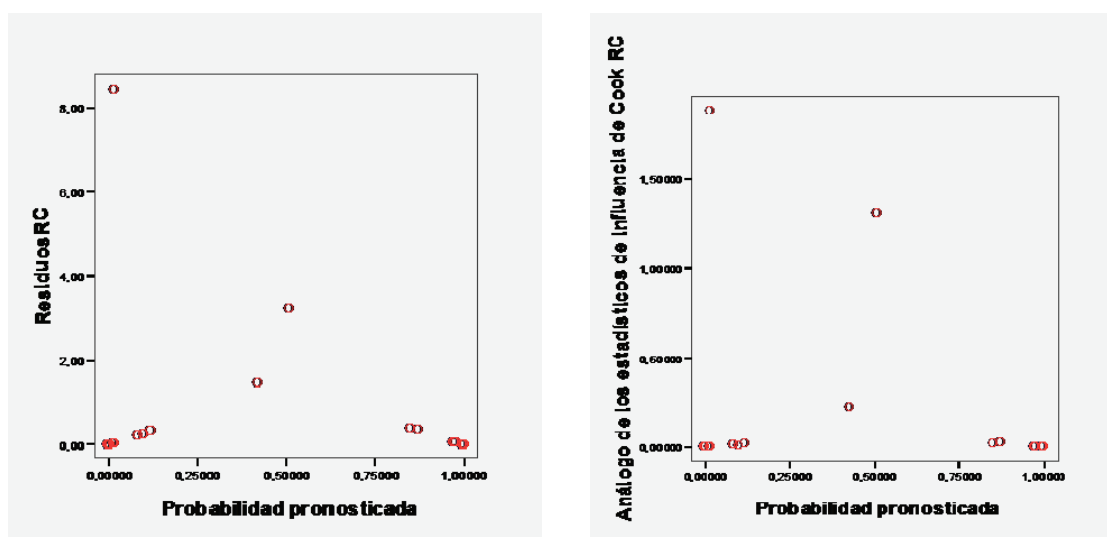


Figura 19. Gráficos de residuos y de influencia para el ítem de RC

En cuanto al estudio de los residuales y valores de influencia puede decirse que el número de casos atípicos (a más de 2 desviaciones típicas) es mínimo (sólo 1 caso),

lo cual se refuerza con el examen de la Figura 19, que muestran muy pocos casos anómalos, apoyando la bondad del ajuste de los datos al modelo.

4. 4. 2. 3. Evaluación

Para la evaluación del cumplimiento del supuesto de linealidad se elaboraron los gráficos correspondientes y los resultados se muestran en las Figuras 20 y 21.

Los resultados obtenidos permiten afirmar que no hay una violación seria del supuesto de linealidad para ninguna de las variables. Como afirma Jovell (1995) es difícil que las relaciones entre variables sean perfectamente lineales y como se observa los gráficos de las Figuras 20 y 21 muestran su aproximación a una línea recta.

La evaluación de la posible existencia de multicolinealidad entre las variables del modelo se hizo a través del examen de la matriz de correlaciones, la cual muestra un valor elevado para los componentes de codificación y combinación (0.813). En razón de este valor de correlación se hizo el diagnóstico de colinealidad, examinando los valores de tolerancia (0.372) y del factor de inflación de la varianza (2.691). También se realizó un análisis de componentes principales con los predictores, según el procedimiento descrito por Hair y cols. (2001), examinando el valor del índice de condicionamiento y después la proporción de varianza explicada por cada componente identificado en el análisis en cada predictor del modelo. Dado que el valor obtenido para el índice de condicionamiento es inferior a 15 y que el valor del factor de inflación de la varianza es considerablemente inferior a la cota habitualmente utilizada de 8 y, sobre todo, dada la naturaleza de las dos variables del análisis, se consideró que la relación existente entre estos dos componentes de procesamiento no representaba un problema serio para la formulación del modelo.

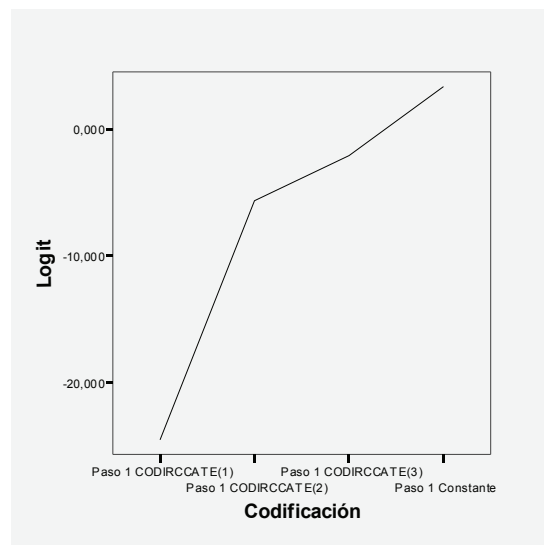


Figura 20. Relación entre el logit estimado y la variable CODI para el formato RC

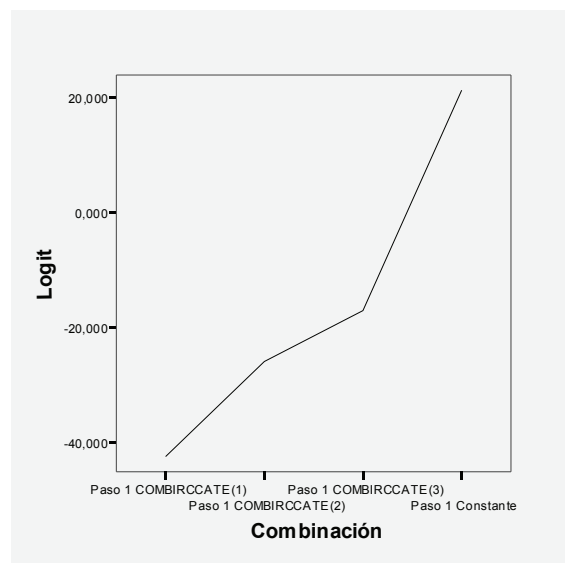


Figura 21. Relación entre el logit estimado y la variable COMB para el formato RC

La evidencia proporcionada por la validación cruzada del modelo también es favorable: la tasa de predicciones correctas en la muestra de validación (99 %) es prácticamente igual a la obtenida al estimar el modelo con todos los sujetos de la muestra (véase Tabla 35).

4. 4. 2. 4. Interpretación

Los resultados obtenidos indican que las variables que ejercen influencia en la respuesta al ítem cuando es formulado con un formato de RC son las correspondientes a las fases de codificación y de combinación de los componentes de procesamiento cognitivo. El resto de variables relativas tanto a metacognición como a percepción sobre la tarea no muestran una influencia importante cuando son consideradas conjuntamente con el resto de variables que se pusieron a prueba.

4. 4. 3. Modelo EA

Dada la naturaleza ordinal de la variable respuesta en el ítem de EA, se comenzó el análisis estimando un modelo de regresión ordinal. No obstante, al violarse en el modelo finalmente obtenido el supuesto de líneas de regresión paralelas en las distintas categorías de respuesta, se optó por recodificar la respuesta (con 0 para respuestas incorrectas o parcialmente correctas y 1 para respuestas correctas) y se estimó un modelo de regresión logística binaria, siguiendo idéntico procedimiento que para los ítems de EM y RC.

4. 4. 3. 1. Especificación

Para el análisis multivariante se incorporó todo el conjunto de variables seleccionadas en base a la teoría, dado que todas ellas mostraron relación significativa con la respuesta de los examinados en el ítem de evaluación de la actuación.

4. 4. 3. 2. Selección

Al trabajar individualmente con los cuatro bloques de variables consideradas en el modelo, los predictores que mostraron un peso significativo fueron:

- a. Conocimiento procedimental, dentro del bloque Conocimiento

- b. Combinación y respuesta, dentro del bloque Componentes de procesamiento
- c. Todas las variables que conforman el bloque metacognitivo: conocimiento autopercebido, planificación, control y conocimiento estratégico
- d. Todas las variables que conforman el bloque relativo a las tareas: valor percibido en relación tanto al contenido como al formato y la dificultad percibida también en relación tanto al contenido como al formato de la pregunta.

Se procedió seguidamente a la estimación del modelo final incorporando las variables seleccionadas de cada bloque, operando de la forma ya indicada, obteniéndose finalmente con los dos métodos utilizados el modelo que se presenta en la Tabla 36, en el cual las variables conocimiento procedimental, combinación, planificación, control y valor percibido de la tarea en relación al formato muestran que hacen un aporte estadísticamente significativo en la explicación de la respuesta al ítem EA, el resto de variables tienen valores para el estadístico de Wald no significativos. Los estadísticos H-L, G y R^2 de Nagelkerke muestran valores que indican que el modelo trabaja satisfactoriamente. El porcentaje de casos correctamente pronosticados revela una excelente capacidad predictiva del modelo finalmente seleccionado.

En cuanto al estudio de los residuales y valores de influencia puede decirse que el número de casos atípicos (a más de 2 desviaciones típicas) es muy pequeño (sólo 6 casos). La Figura 22 muestra muy pocos casos que puedan considerarse anómalos o influenciales, apoyando la bondad del ajuste de los datos al modelo.

Tabla 36. Resultados del modelo de regresión logística para el formato EA

Variables	B	ET	Wald	gl	Sig.	Exp (B)	Límite inferior	Límite superior
CP	0.675	0.161	17.572	1	0.000	1.694	1.433	2.694
COMB	1.343	0.209	41.486	1	0.000	3.831	2.546	5.764
PLAN	1.067	0.478	4.991	1	0.025	2.906	1.140	7.410
CONT	2.252	0.350	41.384	1	0.000	9.511	4.788	18.890
VTF	0.303	0.096	9.999	1	0.002	1.353	1.122	1.633
Constante	-30.966	3.993	60.156	1	0.000	0.000		
Ajuste global								
H-L (sig.)	G (sig.)		R² de Nagelkerke		% CP			
3.408 (0.906)	298.998 (0.000)		0.888		94.7			

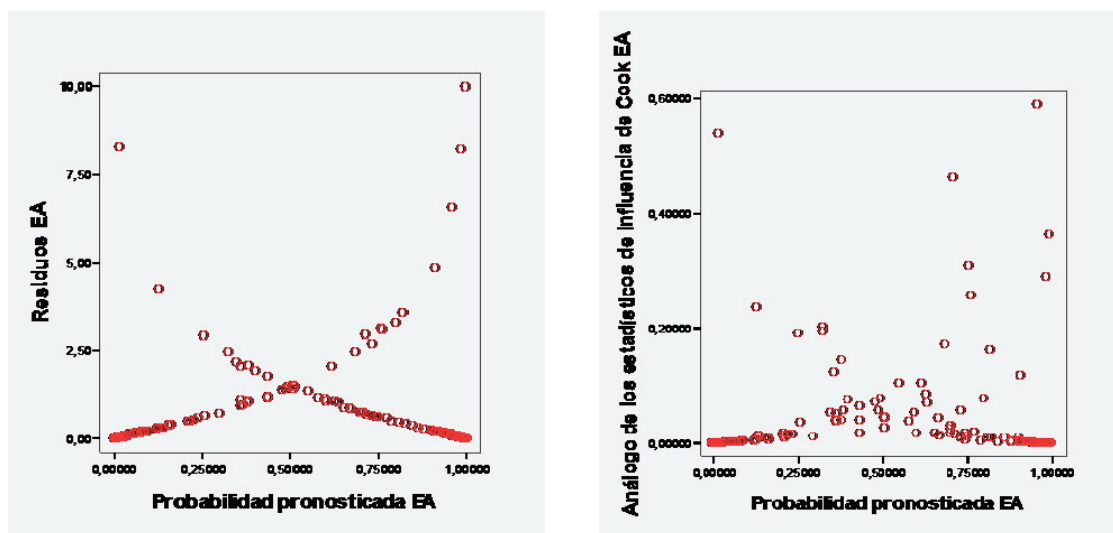


Figura 22. Gráficos de residuos y de influencia para el ítem de EA

4. 4. 3. 3. Evaluación

Para la evaluación del cumplimiento del supuesto de linealidad se elaboraron los gráficos correspondientes y los resultados se muestran en las Figuras 23 a 27.

Los resultados muestran una relación aproximadamente lineal para las variables conocimiento procedimental y combinación y una relación lineal con la variable valor percibido de la tarea en relación al formato. En cuanto a las variables planificación y control se podría afirmar que no hay una violación seria del supuesto de linealidad; aun cuando no se cumple estrictamente, se observa en buena parte del gráfico una relación

lineal bastante clara, junto a una asíntota en la parte superior y/o inferior de la curva. La Figura 28 con los gráficos que muestran la relación entre esas variables y la probabilidad pronosticada de responder correctamente es monótona creciente y es posible suponer en base a esta relación que la desviación del supuesto no ha afectado en gran medida las estimaciones del modelo, ni se han infraestimado las relaciones entre los predictores y la variable respuesta.

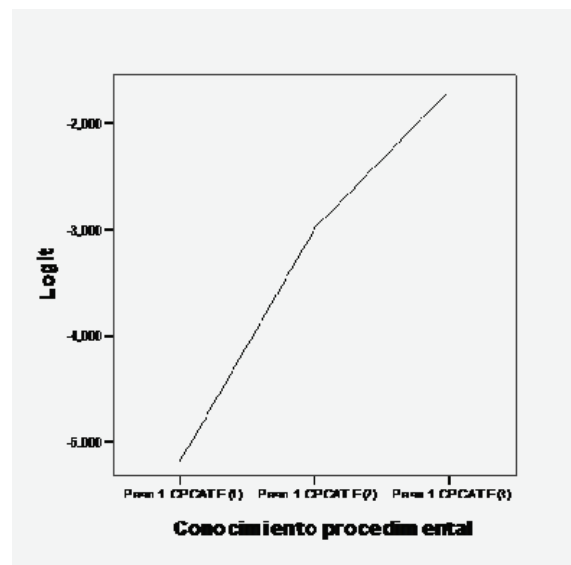


Figura 23. Relación entre el logit estimado y la variable CP para el formato RC

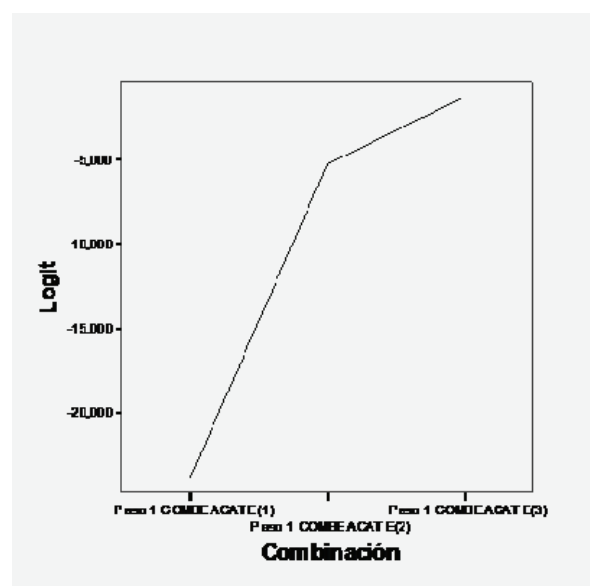


Figura 24. Relación entre el logit estimado y la variable COMB para el formato RC

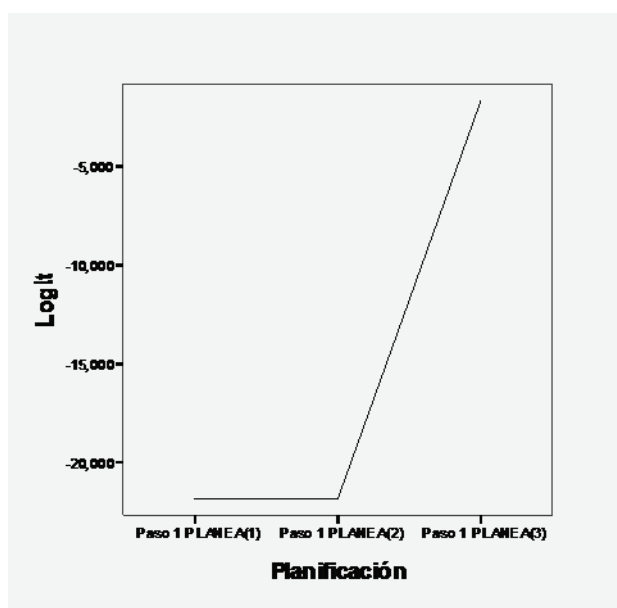


Figura 25. Relación entre el logit estimado y la variable PLAN para el formato RC

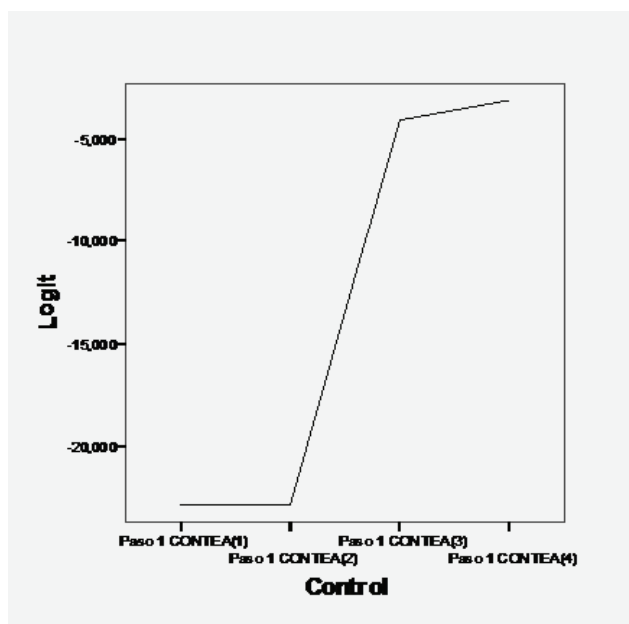


Figura 26. Relación entre el logit estimado y la variable CONT para el formato RC



Figura 27. Relación entre el logit estimado y la variable VTF para el formato RC

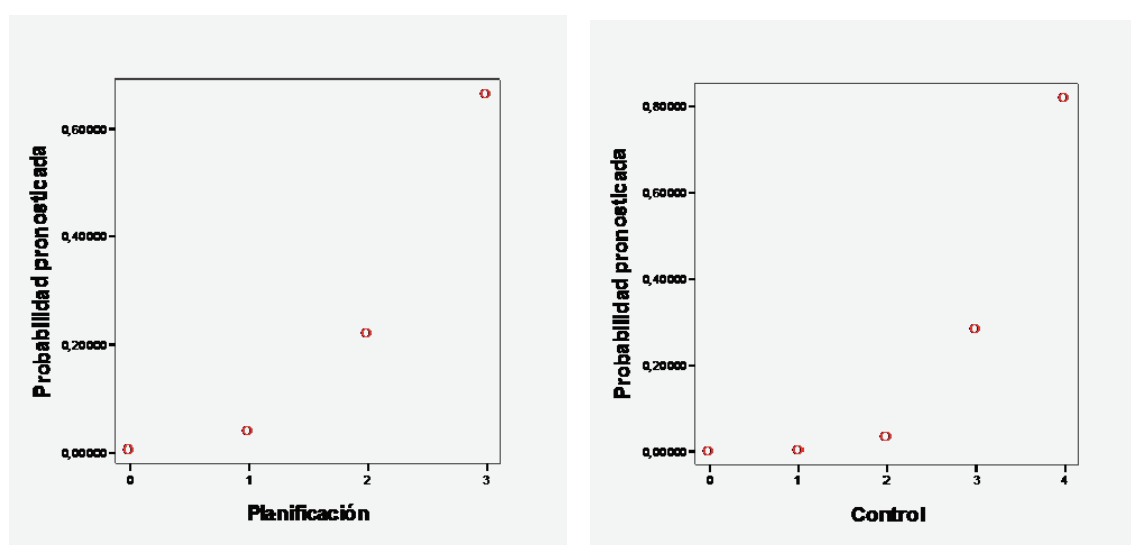


Figura 28. Relación entre las variables PLAN y CONT y la probabilidad de respuesta correcta al ítem con formato EA

La evaluación de la posible existencia de multicolinealidad entre las variables del modelo se hizo a través del examen de la matriz de correlaciones (véase Tabla 37),

la cual no evidencia problemas, dado que no existen valores que puedan suponer riesgo de violación del supuesto (el valor más alto obtenido es 0.621) y en consecuencia, en la estimación del modelo.

Tabla 37. *Correlaciones de las variables incluidas en el modelo para el formato EA*

Variables	CP	COMB	PLAN	CONT
COMB	0.621	-	-	-
PLAN	0.459	0.590	-	-
CONT	0.414	0.594	0.433	-
VTF	0.392	0.527	0.423	0.356

La evidencia proporcionada por la validación cruzada del modelo también es favorable: la tasa de predicciones correctas en la muestra de validación (94.4) es prácticamente idéntica a la obtenida al estimar el modelo con todos los sujetos de la muestra (véase Tabla 36).

4. 4. 3. 4. Interpretación

Los resultados obtenidos indican que las variables que ejercen influencia en la respuesta al ítem cuando es formulado con un formato de EA son el conocimiento procedimental y la fase de combinación del bloque cognitivo, las variables planificación y control del bloque metacognitivo y el valor percibido de la tarea en relación al formato correspondiente al bloque de variables relacionadas con la percepción sobre la tarea.

4. 4. 4. Comparación de los modelos obtenidos en los distintos formatos

En la Tabla 38 se presenta un resumen de los resultados de los modelos de regresión que mejor ajustaron para cada formato estudiado. Como se advierte del examen de esta tabla, la variable combinación se muestra como un predictor común para cualquiera de los tres formatos, informando de la importancia que tienen para el

sujeto la facilidad a la hora de organizar la información, de realizar los cálculos pertinentes y de verificar la adecuación de la solución.

Tabla 38. *Resumen de los resultados para EM, RC y EA*

	FORMATOS DE ÍTEM		
	EM	RC	EA
Variables en el modelo	CD, COMB y CONT	CODI y COMB	CP, COMB, PLAN, CONT y VTF
H-L (p)	0.615 (0.961)	0.058 (1.000)	3.408 (0.906)
G (p)	421.262 (0.000)	560.948 (0.000)	298.998 (0.000)
R ² de Nalgelkerke	0.810	0.989	0.888
% CP	90.7	99.5	94.7

En el estudio de las variables pertenecientes al bloque relativo al conocimiento se observa que la respuesta al ítem de EM es influenciada por el conocimiento declarativo, mientras que la respuesta al ítem de EA lo es por el conocimiento procedimental. En el modelo para el formato de RC ninguna de las variables relativas al conocimiento muestra un aporte que signifique una mejoría importante del modelo.

Por su parte, de las variables del bloque metacognitivo, la actividad de regulación cognitiva control se revela como importante para dar respuesta al ítem cuando es formulado en los formatos de EM y EA. No ocurre lo mismo para el formato de RC, pues aun cuando la variable control había sido seleccionada en el modelado parcial, debido a la fuerza predictiva de las variables codificación y combinación, el efecto de cualquiera de las otras variables queda diluido u oculto.

En cuanto a las variables del bloque relacionado con la percepción sobre la tarea, solamente el modelo para el formato de EA retiene una de ellas, el valor percibido de la tarea considerado en relación al formato en que es presentada. En el modelado parcial,

cuando se trataron independientemente del resto de variables, todas las variables de este bloque resultaron seleccionadas como posibles predictores de la respuesta al ítem, cualesquiera fuera el formato considerado.

Otro detalle interesante es el hecho de que el modelo que más predictores retiene es el correspondiente al formato EA, que corresponde a una pregunta de respuesta abierta, mientras que el que menos predictores retiene es RC, que sólo admite una respuesta corta. Comparando los tres modelos también se observa que el modelo con índices de ajuste y de predicción más altos es el hallado para RC, con un porcentaje de casos correctamente pronosticados muy elevado.

CAPÍTULO V

DISCUSIÓN Y CONCLUSIONES

*There is no true interpretation of anything; interpretation is a vehicle in the service
of human comprehension.*

Andreas Buja

El estudio del formato de los ítems ha sido un tema de creciente interés a lo largo del tiempo, tal y como lo confirman un número importante de publicaciones (véase Rodríguez, 2002, 2003 o Bennett y Ward, 1993). Como ya se ha mencionado, algunos de estos estudios muestran entre sus hallazgos la existencia de diferencias en las propiedades métricas de las puntuaciones obtenidas con ítems de distinto formato, sin llegar a explicar donde residen éstas. En esta investigación se intentó buscar la naturaleza de esas diferencias, para lo cual, partiendo de una exhaustiva revisión bibliográfica, se planteó un modelo que incluye, con una cobertura suficientemente amplia, aquellos constructos que las distintas teorías señalan como responsables de las conductas de respuesta al ítem. El modelo teórico planteado considera un conjunto de variables de distinta naturaleza (cognitiva, metacognitiva y relacionadas con la percepción del examinado sobre la tarea), que podrían ser activadas diferencialmente según el formato en que se plantea la tarea.

El señalamiento de autores como Wilkinson (1999) o Cudeck y Henley (1991), en relación a que un modelo no es capaz de explicar a cabalidad la complejidad de la conducta humana, resulta perfectamente plausible. No obstante, un modelo puede considerarse una aproximación útil para la explicación y síntesis, la cual contribuye a la formalización de esos múltiples procesos y en consecuencia permite un acercamiento al significado de la naturaleza del constructo cuando es medido con diferentes formatos de ítem. En este sentido, el análisis de los resultados obtenidos al someter a contrastación empírica el modelo siempre se ha realizado atendiendo fundamentalmente al significado del modelo, sin desmedro de la atención necesaria a los índices de ajuste y diagnóstico.

Antes de entrar en la discusión de los hallazgos es forzoso destacar que el desarrollo de una investigación de esta naturaleza implica un recorrido no exento de dificultades y obstáculos de distinta índole, que si no son controlados, bien podrían

afectar la posibilidad de conclusiones amplias y generalizables y restringir el alcance de los hallazgos.

El hecho de contar únicamente con un ítem por formato constituye una fuerte limitación en esta investigación, que no permite comparar las propiedades psicométricas de los distintos formatos y en consecuencia restringe la posibilidad de generalización amplia de las conclusiones. Realizar investigaciones con pruebas que contengan un número conveniente de ítems por formato permitiría comparar tanto la equivalencia psicométrica como la equivalencia psicológica, aspiración que no fue posible alcanzar en este estudio por la dificultad que representa una recogida de datos de esta naturaleza, sin contar con los recursos y la colaboración necesarios y suficientes. También debe señalarse que los procesos considerados podrían actuar de otro modo en otros contextos (educación superior o ambientes laborales) u otros países (donde se prefiera un determinado tipo de formato sobre otro) y que deben contemplarse todas estas posibilidades para futuras investigaciones.

No obstante, el rigor seguido en cada uno de los procesos, tanto de revisión de la literatura como de recolección y análisis de datos, permiten tener razonable confianza en los resultados obtenidos, aun cuando se debe reconocer que con mayores recursos podrían afinarse las conclusiones.

De la comprobación empírica del modelo y con los resultados obtenidos, es posible hacer el análisis a dos niveles: uno general, fundado en los resultados del análisis de regresión logística, y uno más particular, basado en los hallazgos parciales encontrados en el camino de la construcción del modelo que dan lugar a otras observaciones que podrían resultar de interés.

En una primera mirada a los modelos de regresión logística obtenidos, se observa que existen coincidencias fundamentales en el comportamiento del sujeto al

responder el ítem, independientemente de su formato. En principio, la fase de combinación de los componentes de procesamiento cognitivo implicados en la tarea es una variable que está presente en los modelos obtenidos con los tres formatos, lo que permite sea considerada como el predictor más importante. Este hallazgo es perfectamente coherente con lo esperado. La fase de combinación constituye el conjunto de procesos cognitivos orientados a reunir y organizar la información, hacer los cálculos necesarios y comparar los resultados obtenidos. En la medida en que este conjunto de procesos resulta más fácil para el examinado es más probable que su respuesta sea la correcta. En este sentido, los resultados apuntan a la necesidad de considerar modelos cognitivos para analizar la ejecución de tareas, identificando los componentes del procesamiento cognitivo involucrados en la resolución de la tarea y estableciendo la forma de darle puntuación a cada componente del modelo (Embretson, 1993).

Continuando el examen de los modelos, los resultados obtenidos también permiten afirmar que las variables de naturaleza metacognitiva constituyen constructos importantes que pueden ayudar a explicar la ejecución, al quedar variables de este bloque en el modelo tanto para el formato de EM como para el de EA. El hecho de que no sean retenidas en el modelo obtenido para el formato RC se debe a la fuerza que tienen los componentes de codificación y combinación en ese formato. Estos resultados son también coherentes con los obtenidos en otras investigaciones (Dowson y McInerney, 2004; Sternberg, 1998a; Pintrich, 2002), que señalan la influencia de los factores metacognitivos y que el formato de respuesta abierta induce a un mayor uso de estrategias de control cognitivo (O'Neil y Brown, 1998), tal como se ha hallado en los datos de este estudio.

En cuanto a las variables relacionadas con la percepción sobre la tarea, su influencia se oscurece completamente debido al marcado efecto de las variables

relacionadas con el sujeto. Solamente el valor percibido de la tarea en relación al formato queda en el modelo para el ítem de evaluación de la actuación.

Ahora bien, no se debe perder de vista que la formulación de este modelo no fue hecha con fines predictivos, sino que su propósito fundamental es identificar las variables que pueden ayudar a explicar el significado de las puntuaciones obtenidas cuando el ítem es medido con distintos formatos, razón por la cual es importante mirar con suficiente detalle los resultados parciales que se desprenden del modelado, así como también todo lo revelado en los análisis preliminares.

Tal como se explicó, el modelado se hizo formato a formato (EM, RC y EA) y por grupos de variables según la dimensión o el ámbito al que pertenecen (cognitivas, metacognitivas y en relación a la tarea). Se comprueba, de esta forma, que la mayoría de las variables de los distintos bloques considerados en el modelo ejercen una influencia sobre la respuesta al ítem cualquiera sea el formato, pero que al ser consideradas de forma conjunta, el efecto de algunas se muestra más importante, diluyendo en algunos casos el efecto del resto de los predictores. También se verifica que el efecto de cualquiera de los factores componentes del modelo es más marcado para los ítems de construcción, en algunos casos para el formato de RC y en otros para el formato de EA, y menos relevante en el ítem con formato de EM.

Se puede observar que el tipo de conocimiento -declarativo o procedimental- que emplean es diferente según el formato afrontado: declarativo para el formato de elección y procedimental para resolver los ítems planteados en formato de respuesta construida. En cuanto a los procesos de ejecución cognitiva, como ya se señalara, el componente que se revela como más importante de todos en cualquiera de los casos es el correspondiente a la fase de combinación, cuya dificultad se valora de modo diferencial en los tres formatos, siendo más difícil para el ítem de respuesta corta y más fácil para el

ítem en formato EM. Esto sucede en los 3 componentes de procesamiento: la dificultad percibida varía con el formato del ítem.

Los modelos por bloques revelan que para los formatos de construcción todas las variables del ámbito metacognitivo son predictores importantes, mientras que para el formato de elección sólo el conocimiento autopercebido y el control lo son. En cuanto al uso de estrategias metacognitivas para la solución de problemas, es importante en este punto observar con detalle lo hallado en los análisis de los datos. En primer lugar se observa que el número de estrategias utilizadas tiene un valor medio superior para la tarea en formato de elección múltiple, pero no existe una relación entre número de estrategias utilizadas y la respuesta al ítem de este formato, relación que sí se observa en los dos formatos de construcción, aun cuando la variable no represente un predictor significativo en el modelo conjunto.

En lo concerniente a las variables relacionadas con la percepción que tiene el sujeto acerca de la tarea se observa que, a pesar de que su efecto desaparece o queda ensombrecido al incorporar variables cognitivas y metacognitivas, resultan predictores significativos tanto en los modelos univariados como en los modelos por bloques. Por otra parte, sin olvidar que el tamaño del efecto en la estimación de las diferencias entre formato para estas variables es muy pequeño, vale destacar que el formato mejor valorado es el de elección múltiple y que los ítems se perciben con mayor dificultad cuando están planteados en un formato de respuesta construida.

Se advierte de lo hallado que las variables estudiadas tienen un comportamiento más similar en los formatos de construcción. Asimismo, se evidencia que la respuesta al ítem, cualesquiera sea su formato, está principalmente determinada por las variables de tipo cognitivo, básicamente por los componentes de ejecución. Aun así, se puede afirmar que todos los constructos estudiados constituyen predictores importantes en la

ejecución del ítem y que la intensidad del efecto de los distintos factores varía de un formato a otro.

A la vista de los resultados obtenidos se puede concluir que cada formato parece demandar del sujeto examinado diferentes conductas, o diferentes niveles en el empleo de su cognición o metacognición. Del mismo modo, la percepción sobre la tarea demandada es diferente según el formato en que se presenta la misma, lo cual, como señalan Pearson y Garavaglia (1997), puede influir en el significado e interpretación de las puntuaciones. Tal como afirma Snow (1993), existe un tejido de elementos que producen diferencias individuales en la respuesta al ítem que tiene relación con el formato.

En consecuencia, para un proceso de medición más integral es recomendable usar tests que incluyan una combinación de formatos, pues cada uno de ellos es capaz de revelar diferencias individuales, no solamente en cuanto a conocimientos y procesos cognitivos, sino también en motivación, esfuerzo y mecanismos de control metacognitivo. Suprimir los ítems de construcción en nombre de la economía y de la precisión puede significar un sacrificio en la capacidad de medición del test y podría dar lugar a inferencias con una validez cuestionable.

En definitiva, la validez y no la conveniencia debe ser la prioridad en la elección del formato del ítem. Dado que cada formato es capaz de elicitar distintas conductas, el formato de respuesta al ítem debería corresponder con la conducta diana en el objetivo de la medición y garantizar que el significado de las puntuaciones obtenidas se corresponda con el constructo que deseamos medir.

REFERENCIAS

- Ackerman, T. y Smith, P. (1988). A comparison of the information provided by essay, multiple-choice and free response. *Applied Psychological Measurement*, 12(2), 117 -128.
- American Psychological Association. (1954). *Technical recommendations for psychological test and diagnostic techniques*. Washington, DC: Autor.
- American Educational Research Association, American Psychological Association y National Council on Measurement in Education. (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- American Educational Research Association, American Psychological Association y National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Anderson, J. (1983). *The architecture of cognition*. Cambridge: Harvard University Press.
- Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.
- Ato, M., Losilla, J. M., Navarro, J. B., Palmer, A. y Rodrigo, M. F. (2005). *Análisis de datos. Modelo lineal generalizado*. Girona: Edicions a Petició, SL.

- Ayala, C., Shavelson, R., Yin, Y. y Shultz, S. (2002). Reasoning dimensions underlying science achievement: the case of performance assessment. *Educational Assessment*, 8, 101-121.
- Ayala, C., Yin, Y., Shultz, S. y Shavelson, R. (2002). *On science achievement from the perspective of different types of test: a multidimensional approach to achievement validation* (Reporte técnico CSE N° 572). California: CRESST/Stanford University.
- Bandura, A. (1990). *Multidimensional scales of perceived academic efficacy*. Stanford, CA: Stanford University.
- Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychology*, 28, 117 – 148.
- Bandura, A. (1995). Exercise of personal and collective efficacy in changing societies. En A. Bandura (ed.), *Self-efficacy in Changing Societies* (pp. 1 - 45). Nueva York: Cambridge University Press.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W. H. Freeman.
- Bandura, A. (2006). Guide for constructing self-efficacy scales. En F. Pajares y T. Urdan (Eds.). *Self-efficacy beliefs of adolescents*, (Vol. 5, pp. 307-337). Greenwich, CT: Information Age Publishing.
- Baxter, G. P. y Glaser, R. (1998). Investigating the cognitive complexity of science achievement. *Educational Measurement: Issues and Practice*. 7(3), 37-45.
- Baxter, G. P., Elder, A. D. y Glaser, R. (1996). Knowledge-base cognition and

- performance assessment in the science classroom. *Educational Psychologist*, 31, 133-140.
- Bejar, I. I. (1985). Speculations on the future of test design. En S. Embretson (Ed.), *Test Design. Developments in Psychology and Psychometrics* (pp. 279 - 294). San Diego, California: Academics Press, Inc.
- Bennett, R. E. (1993). On the meaning of constructed response. En R. E. Bennett y W. Ward. (Eds.), *Construction versus Choice in Cognitive Measurement: issues in constructed response, performance testing and portfolio assessment* (pp. 1 - 27). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Bennett, R. E. y Ward, W. (Eds.) (1993). *Construction versus Choice in Cognitive Measurement: issues in constructed response, performance testing and portfolio assessment*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Bennet, R., Rock, D. y Wang, M. (1991). Equivalence of free-response and multiple-choice items. *Journal of Educational Measurement*, 28, 77 – 92.
- Bennett, R., Morley, M., Quardt, D., Singley, M., Rock, D. Katz, I. et al. (1998, Julio) *Psychometric and cognitive functioning of an underdetermined computer-based response type for quantitative reasoning* (GRE Board Report N° 95-11R). Princeton, NJ: Educational Testing Service.
- Benton, S. L. y Kiewra, K. A. (1987) The assessment of cognitive factors in academia abilities. En R. R. Ronning, J. A. Glover, J. C. Conoley y J. C. Witt (Eds.), *The influence of cognitive psychology on testing*. (pp. 145-189). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

- Bock, R. D., Brennan, R. L. y Muraki, E. (2002). The information in multiple ratings. *Applied Psychological Measurement*, 26, 364-375.
- Borsboom, D. (2005). *Measuring the mind: Conceptual issues in contemporary psychometrics*. Cambridge: Cambridge University Press.
- Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika*, 71, 425-440.
- Borsboom, D. y Mellenbergh, G. J. (2004). Why psychometrics is not pathological: A comment on Michell. *Theory & Psychology*, 14, 105–120.
- Borsboom, D. y Mellenbergh, G. J. (2007). Test validity in cognitive assessment. En J. Leighton y M. Gierl (Eds.), *Cognitive diagnostic assessment for education: Theory and applications*. Cambridge: Cambridge University Press.
- Bourne, L. E., Dominowski, R. L. y Loftus, E. F. (1979). *Cognitive Processes*. Englewood Cliffs, New Jersey: Prentice Hall, Inc.
- Bower, G. H. (1981). Mood and memory. *American Psychologist*, 36, 129-148.
- Brennan, R. L. (Ed.) (2006). *Educational measurement* (4^a ed.). Washington, DC: National Council on Measurement in Education and American.
- Bridgeman, B. (1992). A comparison of quantitative question in open-ended and multiple choice formats. *Journal of Educational Measurement*, 29, 253 – 271.
- Brown, A. (1987). Metacognition, executive control, self-regulation, and other more mysterious mechanisms. En F. E. Weinert y R. H. Kluwe (Eds.), *Metacogniton, motivation, and understanding* (pp. 65 - 116). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

- Bruning, R. H., Schraw, G. J., Norby, M. M. y Ronning, R. R. (2005). *Psicología cognitiva y de la instrucción* (4ª ed.). Madrid: Pearson, Prentice Hall.
- Bunge, M. (1989). *La investigación científica*. Barcelona: Editorial Ariel, S.A.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Nueva York, NY: Cambridge University Press.
- Cattell, R.B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54, 1 - 22.
- Corno, L., Cronbach, L., Kupermintz, H., Lohman, D., Mandinach, E., Porteus, A et al. (2002). *Remaking the concept of aptitude: extending the legacy of Richard E. Snow*. Mahwah, N J: Lawrence Erlbaum Associates, Publishers.
- Cortada de Kohan, N. (2003). Posibilidad de integración de las teorías cognitivas y la psicometría moderna. *Revista Argentina de Neuropsicología*, 1,8 – 23.
- Cronbach, L. J. y Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281- 302.
- Cudeck, R. y Henley, S. J. (1991). Model selection in covariance structures and the “problema” of simple size: A clarification. *Psychological Bulletin*, 109, 512-519.
- Davidson, J. E.; Deuser, R. y Sternberg, R. J. (1994). The role of metacognition in problem solving. En J. Metcalfe y A. P. Shimamura (Eds.), *Metacognition: knowing about knowing* (pp. 207- 226). Massachusetts: Massachusetts Institute of Technology Press.
- De la Torre, J. (2009). A cognitive diagnosis model for cognitively based multiple-

- choice options. *Applied Psychological Measurement*, 33, 163-183.
- Downing, S. M. (2006). Selected-response item formats in test development. En S. M. Downing y T. M. Haladyna (Eds.), *Handbook of test development* (pp. 287-301). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Downing, S. M. y Haladyna, T. M. (2006). *Handbook of test development*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Dowson, M. y McInerney, D. M. (1998, Abril). *Cognitive and motivational determinants of students' academic performance and achievement. Goals, strategies, and academic outcomes in focus*. Documento presentado en el encuentro anual de la American Educational Research Association, San Diego, CA.
- Dowson, M. y McInerney, D. (2004). The development and validation of the goal orientation and learning strategies survey (GOALS-S). *Educational and Psychological Measurement*, 64, 290 -310.
- Dweck, C. y Leggett, E. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, 95, 256-273.
- Eccles, J. y Wigfield, A. (1995). In the mind of the actor: the structure of adolescents' achievement task values and expectancy-related beliefs. *Personality and Social Psychology*, 21, 215-225.
- Eccles, J. y Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual review of psychology*, 53, 109-132.

- Eccles-Parsons, J., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L. et al. (1983). Expectancies, values, and academic behaviours. En J. T. Spence (Ed.), *Achievement and achievement motives: Psychological and sociological approaches* (pp. 75 – 146). San Francisco: Freeman.
- Embretson, S. (1984). A general latent trait model for response processes. *Psychometrika*, 49, 175-186.
- Embretson, S. (Ed.). (1985). *Test design: Developments in psychology and psychometrics*. Nueva York: Academic Press.
- Embretson, S. (1993). Psychometrics models for learning and cognitive processes. En N. Frederiksen, R. Mislevy y I. Bejar (Eds.), *Test theory for a new generation of tests*. (pp. 125– 150). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Embretson, S., y Gorin, J. (2001). Improving construct validity with cognitive psychology principles. *Journal of Educational Measurement*, 38, 343-368.
- Embretson, S. E. y Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Fan, L. y Zhu, Y. (2007). Representation of problem-solving procedures: A comparative look at China, Singapore, and US mathematics textbooks. *Educational studies in mathematics*, 66, 65-71.
- Farley, F. (1985). Psychobiology and cognition: an individual differences model. En J. Strelau, F. Farley y A. Gale (Eds.), *The biological bases of personality and behavior: Vol. 1. Theories, measurement, techniques and development* (pp. 61 - 73). Washington D.C.: Hemisphere

- Ferrara, S., García, T., Freed, R., Vélez-Paschke, A., McGivern, J., Mushlin, S. et al. (2004, abril). Examining test store validity: preliminary analysis of evidence of the alignment of targeted and observed content, skills and cognitive processes in a middle school science assessment. En E. A. Vanderputten (Presidente), *Putting Alignment to the Test Symposium conducted 90 at the annual meeting of the American Educational Research Association*, San Diego, EEUU.
- Fischer, G.H. (1973). The linear logistic test model as an instrument of educational research. *Acta Psychologica*, 37, 359-374.
- Fitzpatrick, C. (1994, abril). *Adolescents mathematical problem solving: the role of metacognition, strategies, and beliefs*. Documento presentado en el encuentro anual de la American Educational Research Association, New Orleans, LA.
- Flavell, J. H. (1992). Metacognition and cognitive monitoring: a new area of cognitive-developmental inquire. En T. O. Nelson (Ed.), *Metacognition. Core Readings* (pp. 3 – 8). Needham Heights, MA: Allyn y Bacon.
- Frederiksen, N., Mislevy, R. J. y Bejar, I. I. (Eds.) (1993). *Tests Theory for a new generation of tests*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- García, T. y McKeachie, W. J. (2005). The making of the Motivated Strategies for Learning Questionnaire. *Educational Psychologist*, 40, 117-128.
- Glaser, R. (1981). The future of testing. A research agenda for cognitive psychology and psychometrics. *American Psychologist*, 36, 923-936.
- Glaser, R., Raghavan, K. y Baxter, G. P. (1992). *Cognitive theory as the basis for design of innovative assessment: Design characteristic of science assessment*

- (Reporte técnico CSE N° 349). Los Angeles: Universidad de California, National Center for Research on Evaluation, Standards, and Student Testing.
- Gorin, J. S. (2006). Test design with cognition in mind. *Educational Measurement: Issues and Practice*, 25(4), 21 – 35.
- Guilford, J. P. (1946). New standards for test evaluation. *Educational and Psychological Measurement*, 6, 427-439.
- Gulliksen, H. (1950/1987). *Theory of mental tests*. Hillsdale: LEA.
- Haertel, E. H. (2006). Reliability. En R. L. Brennan (Ed.), *Educational measurement* (4^a ed., pp. 65–110). Washington, DC: National Council on Measurement in Education and American.
- Haertel, E. H. y Wiley, D. E. (1993). Representations of ability structures. En N. Frederiksen, R. Mislevy y I. Bejar (Eds.), *Test theory for a new generation of tests*. (pp. 359 – 384). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Hair, J., Anderson, R., Tatham, R. y Black, W. (1999). *Análisis Multivariante* (5^a ed.). Madrid: Prentice Hall Iberia
- Haladyna, T. M. (1997). *Writing test items to evaluate higher order thinking*. Needham Heights, MA: Allyn y Bacon.
- Haladyna, T. M. (1999). *Developing and validating multiple-choice test items* (2^a ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Hambleton, R. K. y Zaal, J. N. (Eds.) (1994). *Advances in educational and psychological testing: Theory and applications*. Boston: Kluwer Academic

Publishers.

- Hamilton, L. S., Nusban, E. M. y Snow, R. E. (1997). Interview procedures for validating science assessments. *Applied Measurement in Education*, 10, 181-200.
- Heckhausen, H. (1991). *Motivation and action*. New York, NY: Springer-Verlag.
- Hogan, T. P. (1981). *Relationship between free-response and choice-type test of achievement: a review of the literature* (Reporte N° 70). Washington, DC: National Institute of Education.
- Hogan, T. P. y Agnello, J. (2004). An empirical study of reporting practices concerning measurement validity. *Educational and Psychological Measurement*, 64, 802-812.
- Horn, J. L. (1968). Organization of abilities and the development of intelligence. *Psychological Review*, 75, 242-259.
- Howard, B., Mc. Gee, S., Shia, R. y Hong, N. (2000, abril). *Metacognitive self-regulation and problem solving: expanding the theory base through factor analysis*. Documento presentado en el encuentro anual de la American Educational Research Association, New Orleans, LA.
- Hunt, E. (1987). Science, technology, and intelligence. En R. R. Ronning, J. A. Glover, J. C. Conoley y J. C. Witt (Eds.), *The influence of cognitive psychology on testing* (pp. 11-40). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- INECSE. (2005). *Prueba de matemáticas y de solución de problemas*. Madrid: Instituto

- Nacional de Evaluación y Calidad del Sistema Educativo (INECSE).
- Isen, A., Daubman, K. y Gorgoglione, J. (1987). The influence of positive affect on cognitive organization: Implications for education. En R. Snow y M. Farr (Eds.), *Aptitude, learning and instruction* (Vol. 3, pp. 143-162). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Jackson, D. (1994). *The exploration of a selection of conative constructs relevant to learning and performance*. (Reporte de Proyecto 2.3). Washington, DC: National Center for Research on Evaluating, Standards and Student Testing.
- Janssen, P. y De Boeck, K. (1997). Psychometric modeling of componentially designed synonym task. *Applied Psychological Measurement*, 21, 37-50.
- Jodoin, M. (2003). Measurement efficiency of innovative item format in computer-based testing. *Journal of Educational Measurement*, 40, 1 - 15.
- Jovell, A. J. (1995). *Análisis de regresión logística. Cuadernos metodológicos* (Nº 15). Madrid: Centro de Investigaciones Sociológicas.
- Justice, E. M. y Dornan, T. M. (2001). Metacognitive differences between traditional-age and nontraditional-age college students. *Adult Education Quarterly*, 51, 236-249.
- Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25, 258-272.
- Kane, M. T. (1996). The precision of measurements. *Applied Measurement in*

- Education*, 9, 355-379.
- Kane, M.T. y Mitchell, R. (1996). *Implementing performance assessment. Promises, problems, and challenges*. Mahwah, NJ: LEA.
- Kline, P. (1998). *The new psychometrics: Science, psychology and measurement*. London: Routledge.
- Kluwe, R. H. (1987). Executive decisions and regulation of problem solving behavior. En F. E. Weinert y R. H. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 31-64). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Krantz, D. H., Luce, R. D., Suppes, P. y Tversky, A. (1971). *Foundations of measurement: Vol. I*. New York: Academic Press.
- Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: an overview. *Theory into Practice*, 41(4), 212-218.
- Kuhl, J. (1984). Volitional aspects of achievement motivation and learned helplessness: Toward a comprehensive theory of action control. En B. Maher (Ed.), *Progress in experimental personality research* (Vol. 12, pp. 99-170). Nueva York: Academic Press.
- Kuhl, J. (1985). Volitional mediators of cognition-behavior consistency: Self-regulatory processes and action versus state orientation. En J. Kuhl y J. Beckmann (Eds.), *Action control: From cognition to behavior* (pp. 101-128). Heidelberg, Germany: Springer-Verlag.

- Kuhl, J. y Beckmann, J. (Eds.) (1985). *Action control: From cognition to behavior*. Heidelberg, Germany: Springer-Verlag.
- Kupermintz, H. (2002). Affective and conative factors as aptitude resources in high school science achievement. *Educational Assessment*, 8, 123-137.
- Kyllonen, P. C., Walters, A. M. y Kaufman, J. C. (2005). Noncognitive constructs and their assessment in graduate education: a review. *Educational Assessment*, 10, 153-184.
- Lau, S. y Roeser, R. (2002). Cognitive abilities and motivational processes in high school students' situational engagement and achievement science. *Educational Assessment*, 8, 139-162.
- Leighton, J. P. y Gierl, M. J. (2007). Why cognitive diagnostic assessment? En J. Leighton y M. Gierl (Eds.), *Cognitive diagnostic assessment for education: Theory and applications*. Cambridge: Cambridge University Press.
- Leighton, J. P., Gierl, M. J. y Hunka, S. (2004). The attribute hierarchy model: An approach for integrating cognitive theory with assessment practice. *Journal of Educational Measurement*, 41, 205 -236.
- Linn, R. L. (Ed.) (1989). *Educational measurement*. New York: Macmillan.
- Lohman, D. F. (1993). Teaching and testing to develop fluid abilities. *Educational Researcher*, 22, 12-23.
- Lohman, D. y Ippel, M. (1993). Cognitive Diagnosis: Form statistically bases assessment toward theory-based assessment. En N. Frederiksen, R. J. Mislevy y

- I. I. Bejar (Eds.) *Tests theory for a new generation of tests*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Luce, R. D., Krantz, D. H., Suppes, P. y Tversky, A. (1990). *Foundations of measurement: Vol. III*. New York: Academic Press.
- Maris, E. (1995). Psychometric latent response models. *Psichometrika*, 60, 523-547.
- Martínez Arias, R. (2010). La evaluación del desempeño. *Papeles del Psicólogo*, 31, 85– 96.
- Martinez, M. (1999). Cognition and the question of test item format. *Educational Psychologist*, 34, 207 – 218.
- Marton, F. (1983). Beyond individual differences. *Educational Psychology*, 3, 291-305.
- Masters, G. N. y Mislevy, R.J. (1993). New views of student learning implications for educational measurement. En N. Frederiksen, R. , Mislevy y I. Bejar (Eds.), *Test theory for a new generation of tests* (pp. 219– 242). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers
- Mateos, M. (1995). *Mente y computación. Fundamentos conceptuales de la psicología cognitiva*. Madrid: Ediciones de la Universidad Autónoma de Madrid.
- Mayer, R. E. y Wittrock, M. C. (1996). Problem solving transfer. En D. C. Berliner y R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 47-62). New York, NY: Macmillan.
- McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive Gf-Gc framework. En D. P. Flanagan, J. L. Genshaft y

- P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 151-179). New York: Guildord.
- McNamara, T. P. (1994). Knowledge representation. En R. J. Sternberg (Ed.), *Thinking and problem solving* (pp. 81-117). San Diego, California: Academic Press, Inc.
- Meijer, J., Veenman, M. V. J. y van Hout-Wolters, B. H. (2006). Metacognitive activities in text-studying and problem-solving: development of a taxonomy. *Educational Research and Evaluation*, 12, 209 – 237.
- Messick, S. (1989). Validity. En R.L. Linn (Ed.), *Educational measurement (3ª ed., pp.13-103)*. Nueva York: Macmillan.
- Messick, S. (1993). Construct validity and constructed-response tests. En R. E. Bennett y W. Ward. (Eds.), *Construction versus Choice in Cognitive Measurement: issues in constructed response, performance testing and portfolio assessment* (pp. 61-73). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice*, 14(4), 5-8.
- Messick, S. (1998). Test validity. A matter of consequences. *Social Indicators Research*, 45, 35-44.
- Michell, J. (1999). *Measurement in Psychology. A critical history of a methodological concept*. Cambridge: Cambridge University Press.
- Michell, J. (2000). Normal science, pathological science and psychometrics. *Theory & Psychology*, 10, 639–667.

- Michell, J. (2004). Item response models, pathological science and the shape of error. Reply to Borsboom and Mellenbergh. *Theory & Psychology*, 14, 121–129.
- Mislevy, R. J. (2006). Cognitive psychology and educational assessment. En R. L. Brennan (Ed.), *Educational measurement* (4ª ed., pp. 257-305). Washington, DC: National Council on Measurement in Education and American Council on Education.
- Mullis, I., Martin, M., Ruddock, G., O’Sullivan, C., Arora, A. y Erberber, E. (2005). *TIMSS 2007 Assessment framework*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
- Mumford, M.D., Baughman, W.A., Supinski, E.P. y Anderson, L.E. (1998). A construct approach to skill assessment: procedures for assessing complex cognitive skills. En M. Hakel (Ed.), *Beyond multiple choice: evaluating alternatives to traditional testing for selection* (pp. 75-112). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Muñiz, J. (2004). La validación de los tests. *Metodología de las ciencias del Comportamiento*, 5, 121-141.
- Narens, L. y Luce, R. D. (1986). Measurement: the theory of numerical assignment. *Psychological Bulletin*, 99, 166-80.
- Navas, M. J. (2003). Precisión y análisis de varianza. *Metodología de las Ciencias del Comportamiento*, 5, 73-96.
- Neidorf, T.S., Binkley, M., y Stephens, M. (2006). *Comparing Science Content in the National Assessment of Educational Progress (NAEP) 2000 and Trends in*

- International Mathematics and Science Study (TIMSS) 2003 Assessment* (NCES 2006–026). Washington, DC: National Center for Education Statistics. Recuperado de <http://nces.ed.gov/pubsearch>.
- Newell, A. y Simon, H. A. (1972). *Human Problem Solving*. Englewood Cliffs, N.J: Prentice-Hall, Inc.
- Nickerson, R. S. (1994). The teaching of thinking and problem solving. En R. J. Sternberg (Ed.), *Thinking and problem solving* (pp. 409-449). San Diego, California: Academic Press, Inc.
- Norris, S., Macnab, J. y Phillips, L. (2007). Cognitive modelling of performance on diagnostic achievement test: a philosophical analysis and justification. En J. Leighton y M. Gierl (Eds.), *Cognitive diagnostic assessment for education: Theory and applications*. Cambridge: Cambridge University Press.
- Nunnally, J. C. y Bernstein, I.H. (1995). *Teoría psicométrica*. México: McGraw-Hill.
- OECD. (2009). *PISA 2009 Assessment Framework. Key competencies in reading, mathematics and science*. Recuperado el 05 de septiembre de 2010 de <http://www.oecd.org/dataoecd/11/40/44455820.pdf>.
- O’Neil, Jr., H. F. y Abedi, J. (1996). Reliability and validity of a State Metacognitive Inventory: potential for alternative assessment. *Journal of Educational Research*, 89, 234- 245.
- O’Neil, Jr., H. F. y Brown, R. (1998). Differential effects of question formats in math assessment on metacognition and affect. *Applied Measurement in Education*, 11, 331-351.

- O'Neil, H. F. Jr. y Schacter, J. (1997). *Test specification for problem-solving assessment*. (Reporte técnico CSE N° 463). Los Angeles, CA: National Center for Research on Evaluating, Standards and Student Testing.
- Osterlind, S. (1998). *Constructing test items: multiple-choice, constructed-response, performance, and others formats*. Boston: Kluwer Academia Publishers.
- Paris, S., Lipson, M. y Wixson, K. (1983). Becoming a strategic reader. *Contemporary Educational Psychology*, 8, 293-316.
- Pearson, P. D. y Garavaglia, D. R. (1997). *NAEP Validity Studies: Improving the information value of performance items in large scale assessments. Working paper series* (NAEP Reporte N° NCES-WP-2003-08). Washington, DC: National Center for Education Statistics.
- Pekrun, R. (1994, Abril). *Academic emotions in students' self regulated learning*. Documento presentado en el encuentro de la American Education Research Association, New Orleans.
- Pellegrino, J. W, Chudowsky, N. y Glaser, R. (Eds.) (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.
- Pintrich, P. R. (2002). The role of metacognitive knowledge in learning, teaching, and assessing. *Theory into Practice*, 41, 219-225.
- Pintrich, P. R. y DeGroot, E. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82, 33-40.

- Pintrich, P. R. y Schunk, D. H. (2002). *Motivation in Education .Theory, Research, and Applications. 2nd Edition.* Upper Saddle River, N J: Merrill Prentice Hall.
- Pintrich, P. R., Smith, D. A., García, T. y McKeachie, W.J. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ).* Ann Arbor: University of Michigan, National Center for Research to Improve Postsecondary Teaching and Learning.
- Pintrich, P. R., Smith, D. A., García, T. y McKeachie, W.J. (1993). Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and Psychological Measurement*, 53, 801-813.
- Popper, K. R. (1983). *Conjeturas y refutaciones. El desarrollo del conocimiento científico.* Barcelona: Paidós.
- Prieto, G. (2004). La validez de los tests: situación actual y tendencias futuras. *Metodología de las Ciencias del Comportamiento*, 5, 115-119.
- Prieto, G. y Delgado, A. (1999). Medición cognitiva de las aptitudes. En J. Olea, V. Ponsoda y G. Prieto (Eds.), *Test informatizados: fundamentos y aplicaciones* (pp. 206-226). Madrid: Pirámide.
- Putnam, R. (1987). Mathematics knowledge for understanding and problem solving. *International Journal of Educational Research*, 11, 687-705.
- Rodriguez, M. (2002). Choosing an item format. En G. Tindal y T. Haladyna (Eds.), *Large- scale assessment programs for all students: validity, technical adequacy, and implementation.* (pp. 213-231). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

- Rodriguez, M. (2003). Construct equivalence of multiple-choice and constructed-response items: a random effects synthesis of correlations. *Journal of Educational Measurement*, 40, 163-184.
- Ronning, R. R., Glover, J. A., Conoley, J. C. y Witt, J. (Eds.) (1987). *The influence of cognitive psychology on testing*. Hillsdale, NJ: Lawrence Earlbaum Associates.
- Rupp, A. A., Templin, J. y Henson, R. J. (2010). *Diagnostic measurement: Theory, methods, and applications*. New York: Guilford Press.
- Ryan, A. M. y Greguras G. (1998). Life is not multiple choice: reactions to the alternatives. En M. Hakel (Ed.) *Beyond multiple choice: evaluating alternatives to traditional testing for selection* (pp. 183 - 202). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Savage, L. W. y Ehrlich, P. (1992). *Philosophical and foundational issues in measurement theory*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Schoenfeld, A. H. (1983). Beyond the purely cognitive: belief systems, social cognition, and metacognition as driving forces in intellectual performance. *Cognitive Science*, 7, 329 -363.
- Schoenfeld, A. H. (1985). *Mathematical problem solving*. San Diego, CA: Academic Press.
- Schoenfeld, A.H. (1987). *Cognitive science and mathematics education*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Schraw, G. y Dennison, R. S. (1994). Assessing metacognitive awareness.

- Contemporary Educational Psychology*, 19, 460 - 475.
- Schunk, D. y Zimmerman, B. (Eds.) (1994). *Self-regulation of learning and performance: Issues and educational applications*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Shavelson, R. J. (1972). Some aspects of the correspondence between content structure and cognitive structure in physics instruction. *Journal of Educational Psychology*, 63, 225 - 234.
- Shavelson, R., Ruiz-Primo, M. A., Li, M. y Ayala, C. C. (2003). *Evaluating new approaches to assessing learning*. (Reporte CSE N° 604). California, L A: Center for the Study of Evaluation.
- Shepard, L. A. (1997). The centrality of test use and consequences for test validity. *Educational Measurement: Issues and Practice*, 16(2), 13 - 24.
- Smith, M. U. (Ed.) (1991). *Toward a unified theory of problem solving: views from the contents domains*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Snow, R. E. (1989). Toward assessment of cognitive and conative structures in learning. *Educational Researcher*, 18(9), 8-14.
- Snow, R. E. (1992). Aptitude theory: Yesterday, today, and tomorrow. *Educational Psychologist*, 27, 5-32.
- Snow, R. (1993). Construct validity and constructed-response tests. En R. E. Bennett y W. Ward (Eds.), *Construction versus Choice in Cognitive Measurement: issues*

- in constructed response, performance testing and portfolio assessment* (pp. 45-60). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Snow, R. E. y Farr, M. J. (Eds). (1987). *Aptitude, learning, and instruction: Vol 3. Conative and affective process analyses*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Snow, R. E. y Lohman, D. F. (1989). Implications of cognitive psychology for educational measurement. En R. L. Linn (Ed.), *Educational Measurement* (3^a ed., pp. 263 -332), Nueva York: American Council on Education and Macmillan.
- Snow, R. E. y Lohman, D. F. (1993). Cognitive psychology, new test design, and new test theory: an introduction. En N. Frederiksen, R. Mislevy y I. Bejar (Eds.), *Test theory for a new generation of tests* (pp. 1 – 17). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Snow, R. E. & Peterson, P. L. (1985). Cognitive analyses of tests: implications for redesign. En S. Embretson (Ed.), *Test Design. Developments in Psychology and Psychometrics* (pp. 149 - 166). San Diego, California: Academics Press, Inc.
- Snow, R. E., Corno, L. y Jackson, D., III. (1996). Individual differences in affective and conative functions. En D. C. Berliner y R. C. Calfee (Eds.), *Handbook of Educational Psychology* (243–310). New York: Macmillan.
- Sperling, R. A.; Howard, B. C.; Staley, R. y DuBois, N. (2004). Metacognition and self-regulated learning constructs. *Educational Research and Evaluation*, 10, 117 – 139.

- Sternberg, R. J. (1985). *Beyond IQ. A triarchic theory of human intelligence*. Nueva York: Cambridge University Press.
- Sternberg, R. J. (1994). Cognitive theory and psychometrics. En: R. K. Hambleton y J. N. Zaal (Eds.) *Advances in educational and psychological testing: Theory and applications*. (pp. 367 – 393). Boston: Kluwer Academic Press.
- Sternberg, R. J. (1998a). Abilities are form of developing expertise. *Educational Researcher*, 27(3), 11-20.
- Sternberg, R. J. (1998b). Principles of teaching for successful intelligence. *Educational Psychologist*, 32, 65 - 72.
- Sternberg, R. J., the Rainbow Projects Collaborators y University of Michigan Business School Project Collaborators (2004). Theory-based university admissions testing for a new millennium. *Educational Psychologist*, 39, 185-198.
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103, 677–680.
- Strelau, J. (1983). *Temperament-personality-activity*. Nueva York: Academic Press.
- Sugrue, B. (1993). *Specifications for the design of problem-solving assessment in science*. (Reporte 84117G). Washington, DC: National Center for Research on Evaluating, Standards and Student Testing.
- Sugrue, B. (1995). A theory-based framework for assessing domain-specific problem-solving ability. *Educational Measurement: Issues and Practice*, 14(3), 29-36.
- Suppes, P., Krantz, D. M., Luce, R. D. y Tversky, A. (1989). *Foundations of measurement: Vol. II*. San Diego: Academic Press.

- Traub, R. (1993). On the equivalence of the traits assessed by multiple-choice and constructed-response tests. En: R. E. Bennett y W. Ward. (Eds.), *Construction versus Choice in Cognitive Measurement: issues in constructed response, performance testing and portfolio assessment* (pp. 29-44). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Urdaneta, E y Navas, M. J. (2007, Febrero). *Equivalencia métrica de tres formatos de ítems*. Documento presentado en el X Congreso de Metodología de las Ciencias Sociales y de la Salud, Barcelona, España.
- Watson, D., Clark, L y Tellegen, A. (1988). Development and validation of brief measurement of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54, 1063-1070.
- Whitely, S. E. (1980). Multicomponent latent trait models for ability tests. *Psychometrika*, 45, 479-494.
- Wigfiel, A. (1994). Expectancy-value theory of achievement motivation: a development perspective. *Educational Psychology Review*, 6, 49 -78.
- Willingham, W. (1985). *Success in college*. Nueva York: College Board.
- Wilkinson, L. (1999). *The grammar of graphics*. Springer: Nueva York.

ANEXO A

PRUEBA DE COMPETENCIAS

Apellidos: _____ Nombres: _____
Sexo: _____ Fecha de nacimiento: _____ Liceo: _____
Sección: _____

Lea detenidamente los enunciados de las preguntas y resuelva las cuestiones que se presentan a continuación.

Pregunta A

Este problema trata de cómo organizar el mejor itinerario para unas vacaciones. Las figuras 1 y 2 muestran un mapa de la zona y las distancias en kilómetros entre las ciudades.

Figura 1: Mapa de las carreteras que hay entre ciudades

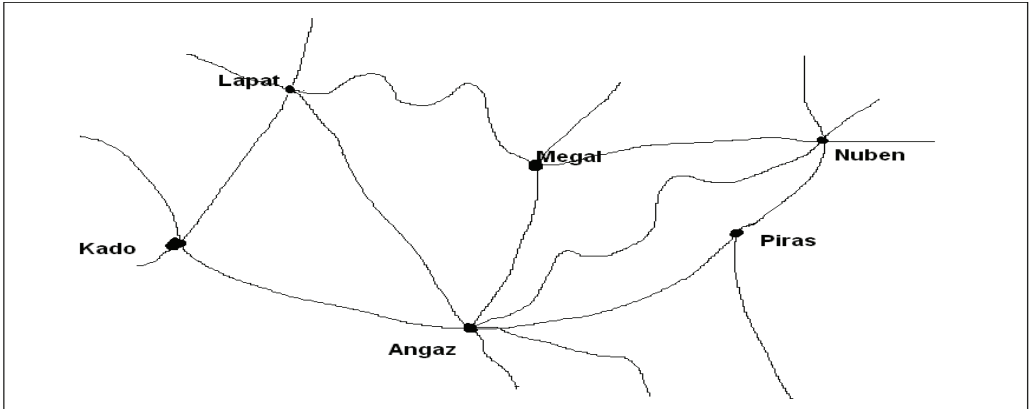


Figura 2: Distancias más cortas entre las ciudades en kilómetros

Angaz						
Kado	450					
Lapat	500	300				
Megal	300	-	550			
Nuben	550	-	-	450		
Piras	250	700	-	-	200	
	Angaz	Kado	Lapat	Megal	Nuben	Piras

Suponiendo que estamos en Nuben, ¿cuál es el recorrido más corto por carretera para llegar a Lapat?

- a. Nuben - Angaz – Lapat
- b. Nuben - Piras- Angaz - Lapat
- c. Nuben - Megal - Lapat
- d. Nuben - Angaz - Kado – Lapat

Pregunta B

Utilizando la información proporcionada en el enunciado de la pregunta anterior, ¿cuál es la distancia más corta por carretera entre Nuben y Kado? Exprese su respuesta en kilómetros.

Pregunta C

El Departamento de Servicios Sociales de Zedlandia está organizando un campamento de cinco días para jóvenes. Se han inscrito en el campamento 46 jóvenes (26 chicas y 20 chicos) y 8 adultos voluntarios (4 hombres y 4 mujeres) atenderán y organizarán el campamento.

ADULTOS
Beatriz
Carolina
Olga
Patricia
Esteban
Ricardo
Guillermo
Pedro

HABITACIONES	
Nombre	Número de camas
Roja	12
Azul	8
Verde	8
Púrpura	8
Naranja	8
Amarilla	6
Blanca	6

Normas de las habitaciones:	
1.	Chicos y chicas deben dormir en habitaciones separadas
2.	Al menos un adulto debe dormir en cada una de las habitaciones
3.	El adulto que duerma en cada habitación debe ser del mismo sexo que el de los jóvenes

Complete la tabla colocando a los 46 jóvenes y a los 8 adultos en las habitaciones según las normas anteriores.

Nombre	Número de chicos	Número de chicas	Nombre del adulto
Roja			
Azul			
Verde			
Púrpura			
Naranja			
Amarilla			
Blanca			

ANEXO B

PRUEBA GENERAL

Apellidos: _____ Nombres: _____

Liceo: _____

A continuación se presentan varios bloques de preguntas para conocer su opinión y su modo de responder a cada uno de los tres problemas anteriores, por lo tanto no hay respuestas correctas o incorrectas.

Lea atentamente cada uno de los enunciados y conteste con el mayor cuidado, exactitud y sinceridad, siguiendo las instrucciones que se le dan. Si hay algo que no entiende o no está muy seguro de cómo responder, puede pedir ayuda.

Siempre debe dar una respuesta en relación a cada una de las tres preguntas A, B y C del cuestionario anterior.

Recuerde que en la pregunta A debía seleccionar entre cuatro opciones de respuesta el itinerario más corto entre dos ciudades, en la pregunta B debía calcular la distancia más corta entre dos ciudades y la pregunta C trataba sobre la organización de las habitaciones en el campamento.

Escriba dentro de cada cuadro el número de la alternativa que mejor expresa su opinión en relación a las preguntas A, B y C, sabiendo que:

1	Completamente en desacuerdo
2	Moderadamente en desacuerdo
3	Moderadamente de acuerdo
4	Completamente de acuerdo

Ejemplo:

La pregunta me pareció difícil

A	B	C
4	3	1

Esta será su respuesta si la pregunta A le pareció muy difícil, la pregunta B un poco difícil y la pregunta C muy fácil.

	A	B	C
1. Creo que contesté correctamente la pregunta	<input type="text"/>	<input type="text"/>	<input type="text"/>
2. Sé lo suficiente como para responder correctamente a la pregunta	<input type="text"/>	<input type="text"/>	<input type="text"/>
3. La tarea planteada en esta pregunta es entretenida	<input type="text"/>	<input type="text"/>	<input type="text"/>
4. Esta pregunta permite demostrar lo que uno sabe	<input type="text"/>	<input type="text"/>	<input type="text"/>
5. La tarea que plantea la pregunta es de poca utilidad para mi formación futura	<input type="text"/>	<input type="text"/>	<input type="text"/>
6. Me gusta la tarea que se plantea en la pregunta	<input type="text"/>	<input type="text"/>	<input type="text"/>
7. La pregunta me pareció difícil	<input type="text"/>	<input type="text"/>	<input type="text"/>
8. Para contestar correctamente esta pregunta es necesario saber mucho	<input type="text"/>	<input type="text"/>	<input type="text"/>

Escriba dentro de cada cuadro el número de la alternativa que mejor expresa su opinión en relación al formato utilizado en la pregunta A (elección), B (respuesta corta) y C (respuesta larga), sabiendo que:

1	Completamente en desacuerdo
2	Moderadamente en desacuerdo
3	Moderadamente de acuerdo
4	Completamente de acuerdo

	A	B	C
9. Me resulta sencillo responder a preguntas con este tipo de formato	<input type="text"/>	<input type="text"/>	<input type="text"/>
10. Las preguntas planteadas con este formato resultan tediosas	<input type="text"/>	<input type="text"/>	<input type="text"/>
11. Es importante saber responder a preguntas formuladas de este modo	<input type="text"/>	<input type="text"/>	<input type="text"/>
12. Si pudiera elegir un tipo particular de examen, me gustaría que las preguntas tuvieran este formato	<input type="text"/>	<input type="text"/>	<input type="text"/>
13. Estar familiarizado con preguntas planteadas con este tipo de formato puede ser de utilidad en el futuro	<input type="text"/>	<input type="text"/>	<input type="text"/>

Cuadernillo 1

14. Con preguntas planteadas con este formato es fácil aprobar un examen

--	--	--

15. Las preguntas planteadas con este formato exigen mucho esfuerzo para ser contestadas correctamente

--	--	--

16. El tipo de formato utilizado hace que la pregunta parezca más difícil

--	--	--

Marque con una equis la opción que refleja la forma en la que ha trabajado al responder a las preguntas A, B y C.

ANTES DE RESPONDER A LA PREGUNTA:

17. ¿ha leído detenidamente la pregunta?

A	B	C						
<table border="1" style="display: inline-table;"><tr><td>Si</td><td>No</td></tr></table>	Si	No	<table border="1" style="display: inline-table;"><tr><td>Si</td><td>No</td></tr></table>	Si	No	<table border="1" style="display: inline-table;"><tr><td>Si</td><td>No</td></tr></table>	Si	No
Si	No							
Si	No							
Si	No							

18. ¿se ha asegurado de haber entendido qué es lo que se pedía?

Si	No	Si	No
----	----	----	----

19. ¿ha reformulado el enunciado con sus propias palabras para poder comprender mejor lo que se pedía?

Si	No	Si	No
----	----	----	----

20. ¿ha buscado y entresacado la información principal proporcionada por el enunciado?

Si	No	Si	No
----	----	----	----

21. ¿ha hecho un pequeño plano de situación con la información proporcionada y con lo que se pedía?

Si	No	Si	No
----	----	----	----

22. ¿se ha planteado alguna estrategia concreta para realizar la tarea requerida?

Si	No	Si	No
----	----	----	----

23. ¿sabía perfectamente lo que tenía que hacer para llegar a la respuesta correcta?

Si	No	Si	No
----	----	----	----

CONFORME IBA RESPONDIENDO A LA PREGUNTA:

24. ¿ha ido chequeando su trabajo?

A	B	C						
<table border="1" style="display: inline-table;"><tr><td>Si</td><td>No</td></tr></table>	Si	No	<table border="1" style="display: inline-table;"><tr><td>Si</td><td>No</td></tr></table>	Si	No	<table border="1" style="display: inline-table;"><tr><td>Si</td><td>No</td></tr></table>	Si	No
Si	No							
Si	No							
Si	No							

25. ¿ha encontrado alguna dificultad para poder realizar la tarea planteada?

Si	No	Si	No
----	----	----	----

*Si contestó NO a la pregunta 25, pase directamente a la pregunta 29.
Si contestó SI a la pregunta 25 responda a la siguiente pregunta.*

¿QUÉ HA HECHO PARA INTENTAR RESOLVER ESA DIFICULTAD?

26. Seguir adelante como si nada

A	B	C						
<table border="1" style="display: inline-table;"><tr><td>Si</td><td>No</td></tr></table>	Si	No	<table border="1" style="display: inline-table;"><tr><td>Si</td><td>No</td></tr></table>	Si	No	<table border="1" style="display: inline-table;"><tr><td>Si</td><td>No</td></tr></table>	Si	No
Si	No							
Si	No							
Si	No							

27. Parar y leer de nuevo el enunciado de la pregunta

Si	No	Si	No
----	----	----	----

28. Buscar una nueva manera de resolver la pregunta

Si	No	Si	No
----	----	----	----

Indique en qué pregunta(s) -A, B, C- utilizó las estrategias que se mencionan a continuación para resolver la tarea planteada. Para ello, marque con una equis la letra que identifica a la pregunta en la que haya utilizado esa estrategia.

Ejemplo:

¿Trabajó con una estrategia que ha utilizado en otras ocasiones en tareas similares?

A	B	C
---	---	---

Esta será su respuesta si empleó una estrategia que había utilizado antes para responder las preguntas A y C.

29. ¿Consideró varias formas de trabajo para responder a la pregunta?
30. ¿Ha tratado de dividir el problema en pequeños pasos?
31. ¿Planteó una situación similar para comprender la pregunta formulada y poder responderla?
32. Antes de hacer cálculos, ¿descartó las alternativas que no le parecían respuestas posibles?
33. ¿Utilizó procedimientos que, aunque no conducían directamente a la solución, le ayudaron a obtener la respuesta?
34. ¿Respondió la pregunta partiendo de una respuesta probable?
35. ¿Trabajó con una estrategia que ha utilizado en otras ocasiones en tareas similares?
36. ¿Su forma de trabajo ha sido guiada por el análisis lógico del problema planteado en la pregunta?
37. Una vez respondida la pregunta, ¿revisó su trabajo para asegurarse de no haber cometido errores?

Escriba dentro de cada cuadro el número de la alternativa que mejor expresa la dificultad que ha encontrado al realizar las siguientes tareas cuando respondía a las preguntas A, B y C:

1	Muy difícil
2	Difícil
3	Fácil
4	Muy fácil
NP	No Procede

Ejemplo

Hacer cálculos

A	B	C
<input type="text" value="3"/>	<input type="text" value="4"/>	<input type="text" value="NP"/>

Esta será su respuesta si los cálculos para la pregunta A le resultaron fáciles, los de la pregunta B muy fáciles y no realizó cálculos en la pregunta C

- | | | | |
|--|----------------------|----------------------|----------------------|
| 38. Comprender con claridad lo que pide el problema | <input type="text"/> | <input type="text"/> | <input type="text"/> |
| 39. Identificar las variables implicadas en la solución | <input type="text"/> | <input type="text"/> | <input type="text"/> |
| 40. Identificar las condiciones que debe cumplir la solución | <input type="text"/> | <input type="text"/> | <input type="text"/> |
| 41. Organizar la información proporcionada | <input type="text"/> | <input type="text"/> | <input type="text"/> |
| 42. Hacer cálculos | <input type="text"/> | <input type="text"/> | <input type="text"/> |
| 43. Comprobar la adecuación de la solución | <input type="text"/> | <input type="text"/> | <input type="text"/> |
| 44. Seleccionar o escribir la respuesta correcta | <input type="text"/> | <input type="text"/> | <input type="text"/> |

ANEXO C

EJEMPLO DE HOJA DE REGISTRO PARA LOS PROFESORES

EJEMPLO DE HOJA DE REGISTRO

Liceo Bolivariano Cristóbal Mendoza 4° de Cs. B			
Apellidos y nombres	Conocimiento declarativo	Conocimiento procedimental	Calificación en matemáticas
RUZZA SUAN DENYS			
SEGOVIA MÉNDEZ GERARDO			
CARREÑO BRICEÑO CLAUDIA			
NUÑEZ LINARES MIGUEL			
FAJARDO OCANDO GABRIEL			
LINARES OLIVAR GREGORY			
PAREDES LINARES YONNY			
YÉPEZ LINARES BÁRBARA			
TERÁN ALDANA MARÍAN			
MILLA BARRETO ROMELANY			
BRICEÑO DABOÍN YERALDIN			
CASTELLANOS RAFAEL			
PEREZ SUAREZ KAREN			
LINARES BENDERSON			
GONZÁLEZ BRICEÑO LIZ			
RUZA VÁSQUEZ NATALY			
CEDEÑO OLMOS FREDDY			
ALDANA RODRÍGUEZ DANUVIS			
RAGA VÁSQUEZ ELIANA			
ALDANA BRICEÑO YUNIAR			
ROSARIO BARRETO JUANA			
ROSARIO BARROETA JOSÉ			

ANEXO D

MODELO DE CARTA UTILIZADA CON LOS CENTRO ESCOLARES

RESUMEN DEL PROYECTO DE INVESTIGACIÓN

MODELO DE CARTA UTILIZADA CON LOS CENTRO ESCOLARES

Lugar, fecha

Estimado Sr.:
Nombre del director
 Director del *nombre del instituto educativo*
 Presente.-

Actualmente estamos desarrollando una investigación que trata sobre el tema de los formatos utilizados para formular las preguntas en los exámenes de evaluación del rendimiento académico.

A partir de resultados preliminares de la investigación se ha propuesto un modelo y para poner a prueba el mismo es necesario contar con un conjunto de datos relativos a los estudiantes cursantes de 4º de bachillerato, quienes han sido seleccionados como la población objetivo en esta investigación.

En el documento anexo a la presente se describe con algo más de detalle el trabajo en cuestión. No obstante, si lo desea, antes de tomar una decisión un miembro del equipo de investigación puede ponerse nuevamente en contacto con usted para comentar cualquier aspecto del estudio sobre el que desee recabar información adicional.

La colaboración que necesitamos de su centro para poder llevar a cabo la recogida de datos en esta investigación es la siguiente:

- Necesitamos aplicar a todos los estudiantes de 4º de bachillerato un cuestionario de una hora de duración: el cuestionario incluye 3 preguntas de contenido académico cada una con un formato diferente y el resto de preguntas del cuestionario hacen referencia a la forma en la que el estudiante ha respondido a las tres preguntas anteriores. La aplicación de la prueba será realizada por un miembro del equipo de investigación durante el horario escolar.
- Necesitamos que el profesor de matemáticas de estos alumnos realice una valoración cuantitativa sobre dos aspectos relativos al nivel de conocimientos de sus alumnos en esa materia y que nos proporcione la calificación obtenida en la última evaluación realizada en el curso.

Para que esto pueda coordinarse con mayor eficacia consideramos oportuno que se designe una persona de contacto que sirva como enlace entre el encargado de la investigación y el instituto con el propósito de canalizar el flujo de información de modo efectivo.

Le agradecemos altamente su atención y toda la colaboración que pueda prestarnos para llevar a buen destino la investigación pues las conclusiones que se extraigan de la misma sin duda contribuirán al avance y desarrollo de los métodos de evaluación educativa.

Reciba un saludo muy cordial de,

Elsy Urdaneta
 Investigador Principal
 ULA- NURR

RESUMEN DEL PROYECTO PARA CENTROS ESCOLARES

EL FORMATO DE LA PREGUNTA Y LA RESPUESTA AL ÍTEM: UN ESTUDIO PSICOMÉTRICO DESDE LA PERSPECTIVA DE LA PSICOLOGÍA COGNITIVA

La evaluación de los resultados obtenidos en el proceso de enseñanza aprendizaje representa una fase de suma importancia dentro de este proceso, de manera que ésta debe hacerse con los instrumentos más sólidos y adecuados. En este sentido el formato de la pregunta que se utiliza en los exámenes es muy importante, pues existe duda sobre si los distintos tipos de formatos logran medir justamente aquello que se desea evaluar.

Esta es la razón por la cual la selección del formato de la pregunta ha motivado numerosas investigaciones que intentan averiguar si miden lo mismo aquellas preguntas en las que el examinado debe seleccionar una respuesta entre varias o aquellas en las que el estudiante debe elaborar su respuesta, ya sea ésta una simple palabra o frase, o una respuesta larga. Sin embargo, no existe consenso en estas investigaciones empíricas que nos permita dar una respuesta concluyente en este momento.

Todo esto nos conduce a intentar explorar cuáles son los factores que intervienen en la resolución de una pregunta, mediante un modelo que permita estimar la contribución de variables -cognitivas y no cognitivas- que podrían dar cuenta de las diferencias individuales en el rendimiento académico medido a través de las puntuaciones obtenidas por los sujetos en preguntas formuladas con distinto formato.

Objetivo

Proponer un modelo que permita identificar las variables que explican las diferencias individuales en la ejecución de una pregunta, y así determinar si los factores que dan cuenta de la puntuación del estudiante en un examen se expresan diferencialmente en función del formato de pregunta utilizado.

Método

Sujetos

Estudiantes de 4º de bachillerato

Recogida de datos

Estudiantes: aplicación colectiva de un cuestionario de aproximadamente una hora de duración al conjunto de alumnos seleccionados dentro de la clase y aplicación individual utilizando la técnica de pensar en voz alta a un número muy reducido de alumnos.

Profesor: valoración cuantitativa sobre dos aspectos relativos al nivel de conocimientos del alumno en la asignatura matemáticas, así como la calificación obtenida en la última evaluación realizada en el curso.

Instrumentos

El cuestionario que ha de responder el estudiante consta de dos partes: una pequeña prueba de conocimientos con tres preguntas planteadas cada una con un formato diferente y el resto del cuestionario presenta bloques de preguntas para medir las variables consideradas en el modelo que se desea contrastar y que tienen que ver con la forma en la que el alumno ha respondido a las tres preguntas de la primera parte.

ANEXO E

GUIÓN PARA EL APLICADOR DEL TEST

GUIÓN PARA EL APLICADOR DEL TEST

Una vez que todos los alumnos están ubicados en sus asientos, antes de empezar a repartir los cuestionarios nos dirigimos a los alumnos y les decimos:

- Han sido seleccionados para formar parte de una investigación sobre formatos de respuesta y debido a la importancia de la misma les pedimos lo hagan lo mejor que puedan.
- A continuación voy a colocar en sus pupitres un par de cuestionarios a los que no podrán dar vuelta hasta tanto se les de la indicación.

Repartimos los cuestionarios, colocándolos con el anverso sobre el pupitre, recordando a los estudiantes que no se dé la vuelta al material hasta que se le indique:

- Por favor, no den vuelta al cuestionario todavía.

Una vez terminados de repartir los cuestionarios se les indica a los estudiantes que:

- Pueden dar la vuelta al primer cuestionario, por favor escriban sus datos en los espacios correspondientes.
- En este cuestionario encontrarán tres preguntas que no están relacionadas directamente con ninguna de sus asignaturas, de manera que no necesitan haberse preparado para esta prueba en particular, sino que tendrán que hacer uso de los conocimientos y destrezas que han ido acumulando a lo largo de su formación.
- Lean detenidamente los enunciados de las preguntas y resuelvan las cuestiones que se presentan a continuación.
- Tienen 30 minutos para responder este cuestionario, si terminan antes por favor dejen el cuestionario en el pupitre hasta que transcurra este tiempo, cuando se pasará a recogerlos en el pupitre

Aviso 10 minutos antes de expirar el tiempo.

- Tienen 10 minutos para finalizar

Aviso 5 minutos antes de expirar el tiempo.

- Tienen 5 minutos para finalizar

Pasados los 30 minutos se comenzará a recoger los cuestionarios una vez se les pida a los estudiantes que:

- Den vuelta al cuestionario y déjenlo sobre el pupitre mientras se recogen. Por favor no den vuelta al segundo cuestionario hasta recibir la orden.

Una vez recogidos todos los cuestionarios se les pide que den la vuelta al segundo cuestionario, rellenen de nuevo los datos personales y se lee con ellos en voz alta las instrucciones, diciéndoles:

- Ahora den vuelta al segundo cuestionario y coloquen sus datos donde corresponde. En este cuestionario se presentan varios bloques de preguntas para conocer su opinión y su modo de responder a cada uno de los tres problemas anteriores, por lo tanto no hay respuestas correctas o incorrectas.
- Lean atentamente cada uno de los enunciados y contesten con el mayor cuidado, exactitud y sinceridad, siguiendo las instrucciones que se les dan. Si hay algo que no entienden o no están muy seguros de cómo responder, pueden pedir ayuda.
- Siempre deben dar una respuesta en relación a cada una de las tres preguntas A, B y C del cuestionario anterior.

Se termina la lectura de las instrucciones hasta el ejemplo de la primera parte y se les dice cuanto tiempo tiene para responder el cuestionario

- Tienen 35 minutos para responder este cuestionario, pero en caso de que terminen antes pueden levantar la mano para entregarlo.

Aviso 5 minutos antes de expirar el tiempo.

- Tienen 5 minutos para finalizar

Pasado el tiempo acordado se les pide que:

- Den vuelta al cuestionario y déjenlo sobre el pupitre mientras se recogen

Una vez recogidos todos los cuestionarios se les agradece la participación

- Muchas gracias por la colaboración que han prestado al participar en esta prueba

ANEXO F

MODELO DE CARTA PARA LOS PROFESORES DE MATEMÁTICAS

CUADRO EXPLICATIVO SOBRE TIPOS DE CONOCIMIENTO

MODELO DE CARTA PARA LOS PROFESORES DE MATEMÁTICAS

Estimado profesor:

Como seguramente sabe hemos realizado en este instituto un proceso de recogida de datos que servirán para el desarrollo de un proyecto de investigación, del que le adjunto un pequeño resumen. El instrumento utilizado para este fin fue un cuestionario aplicado a algunos de sus estudiantes. Para completar el trabajo de recogida de datos necesitamos su valiosa colaboración, que consiste en proporcionar la calificación obtenida por esos estudiantes en Matemáticas, así como realizar una valoración cuantitativa sobre aspectos relativos a su capacidad para resolver problemas.

En la hoja adjunta encontrará la relación de los alumnos para los que necesitamos disponer de esa información, así como las preguntas que los estudiantes respondieron en el cuestionario. En la primera columna (CALIFICACIÓN) debe introducir la calificación numérica obtenida por estos alumnos en la última evaluación de Matemáticas realizada en el curso. En las columnas segunda y tercera debe valorar en una escala de 1 a 10 el nivel de conocimiento declarativo y procedimental (véase tabla adjunta para una breve caracterización de los mismos) que considere que dispone cada uno de los alumnos para resolver las tareas que plantean esas tres preguntas.

Le rogamos que, en cuanto le sea posible, dedique unos minutos a rellenar dicha tabla, porque esa información es **imprescindible** para poder someter a prueba el modelo que queremos validar en esta investigación, acerca de cuáles son los factores que sirven para explicar la forma en la que un alumno responde a una pregunta de un examen o test.

Dentro de una semana una persona encargada para tal fin pasará por el centro escolar a recoger la información requerida, o si lo prefiere nos la puede hacer llegar por correo electrónico a proyecto.items@psi.uned.es

Altamente agradecida de su colaboración en el proyecto, quedo a sus gratas órdenes,

Elsy J. Urdaneta Durán
Investigador Principal

CUADRO EXPLICATIVO SOBRE TIPOS DE CONOCIMIENTO

TIPO DE CONOCIMIENTO	EN UNA PALABRA	DEFINICIÓN	EJEMPLOS
DECLARATIVO	Saber qué	<p>Conocimiento de términos, hechos específicos, conceptos, relaciones, clasificaciones, leyes y teorías.</p> <p>Va desde elementos básicos (el significado o características perceptivas de las cosas) hasta conocimiento conceptual altamente organizado (comprensión del significado práctico de una ley o teoría).</p>	<p>Saber que un número es mayor o menor que otro</p> <p>Saber que un triángulo equilátero tiene sus tres lados iguales</p> <p>Conocer el significado de un número fraccionario</p> <p>Comprender el significado de la primera ley de la termodinámica</p>
PROCEDIMENTAL	Saber cómo	Conocimiento acerca de cómo hacer algo, aplicar métodos, técnicas, algoritmos o procedimientos para alcanzar una respuesta o solución.	<p>Resolver una ecuación de segundo grado</p> <p>Sumar fracciones</p> <p>Calcular el área de una circunferencia</p>

